# Gradient test under non-parametric random effects models

**Antonio Hermes Marques da Silva Júnior**

A Thesis presented for the degree of
Doctor of Philosophy


Statistics and Probability Research Group
Department of Mathematical Sciences
University of Durham
England

May 2018

*Dedicated to*

to my beloved parents and my family.

# Gradient test under non-parametric random effects models

**Antonio Hermes Marques da Silva Junior**

Submitted for the degree of Doctor of Philosophy
September 2017

## Abstract

The gradient test proposed by Terrell (2002) is an alternative to the likelihood ratio, Wald and Rao tests. The gradient statistic is the result of the inner product of two vectors — the *gradient* of the likelihood under null hypothesis (hence the name) and the result of the difference between the estimate under alternative hypothesis and the estimate under null hypothesis. Therefore the gradient statistic is computationally less expensive than Wald and Rao statistics as it does not require matrix operations in its formula. Under some regularity conditions, the gradient statistic has $\chi^2$ distribution under null hypothesis. The generalised linear model (GLM) introduced by Nelder & Wedderburn (1972) is one of the most important classes of statistical models. It incorporates the classical regression modelling and analysis of variance either for continuous response and categorical response variables under the exponential family. The random effects model extends the standard GLM for situations where the model does not describe appropriately the variability in the data (*overdispersion*) (Aitkin, 1996a). We propose a new unified notation for GLM with random effects and the gradient statistic formula for testing fixed effects parameters on these models. We also develop the Fisher information formulae used to obtain the Rao and Wald statistics. Our main interest in this thesis is to investigate the finite sample performance of the gradient test on generalised linear models with random effects. For this we propose and extensive simulation experiment to study the type I error and the local power of the gradient test using the methodology developed by Peers (1971) and Hayakawa (1975). We also compare the local power of the test with the local power of the tests of the likelihood ratio, of Wald and Rao tests.

**Keywords**: asymptotic test, overdispersion, generalised linear models

# Declaration

The work in this thesis is based on research carried out at the Statistics and Probability Research Group, the Department of Mathematical Sciences, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

# Acknowledgements

The development of this thesis would not have been possible without the assistance of several people, and it is impossible to acknowledge properly every individual contribution.

First and foremost I would like to express my sincere thanks to my main supervisor, Dr Jochen Einbeck, for his encouragement, interest and patience. Personally, I would like to thank him for sharing his knowledge which has enriched my study in statistics. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my queries and questions so promptly and so fondly. His commitment and encouragement allowed me to present our work to local and international conferences, where I became acquainted with the latest and greatest research and met many prestigious scientists.

I sincerely acknowledge my second supervisor Prof Peter Craig due to his continuous support, his kindness and always helpful and brilliant insights. His friendly guidance and expert advice have been invaluable throughout all stages of the work.

I would like to take this opportunity to thank Prof Andrew Wood and Prof Michael Goldstein – my viva examiners, for their very helpful comments and suggestions.

I would also wish to express my gratitude to Prof John Hinde for receiving me at NUI Galway, Ireland and valuable discussions which have contributed greatly to the improvement of this thesis.

I wish to thank to dozens of people at Department of Statistics of Universidade Federal do Rio Grande do Norte in Brazil who immensely helped me before and during my PhD in UK. A special acknowledgement goes to my friends Dr André Pinho and Dr Carla Vivacqua who encouraged and support me before and through my PhD. Another special thanks goes to Dr Damião da Silva who firstly encouraged

me to pursuit a degree outside Brazil.

I would like to acknowledge the financial support through the Brazil's Science Without Borders Program grant nº 9622/13-6 from CAPES foundation.

Last but not least my parents Hermes and Evandir deserve my whole gratitude for their immensely patience in raising me and for providing me the best care and love any parent could provide. I am deeply grateful to my wife Natália and my daughter Clarice for their immeasurable support, love and for the very happy moments.

**May 30, 2018**

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

The *gradient test* is a relatively new asymptotic test proposed by Terrell (2002) as an alternative to the likelihood ratio, Wald and Rao tests. The *gradient statistic* is the inner product between two vectors — the gradient of the log-likelihood under null hypothesis (hence the name) and the difference between the estimate under alternative hypothesis and the null hypothesis. Therefore, the gradient statistic does not have any matrix or matricial operations in its formula, differently from the Wald and Rao statistics. This turns to be the most appealing advantage of the gradient statistic, making it computationally less expensive than the aforementioned tests. The gradient statistic also is approximately chi-squared for sufficiently large sample sizes and under some regularity conditions.

Since then, researchers have explored the finite sample properties of the gradient test for several statistical models. Lemonte & Ferrari (2011b) studied the size and power in Birnbaum–Saunders regression model, Lemonte & Ferrari (2011c) studied testing hypotheses in the Birnbaum–Saunders distribution under type-II censored samples, Lemonte & Ferrari (2011a) evaluated the local power of some tests in exponential family nonlinear models, Lemonte (2012) studied the local power properties of some asymptotic tests in symmetric linear regression models, Lemonte & Ferrari (2012) examined the local power and size properties of the LR, Wald, score and gradient tests in dispersion models, Vargas et al. (2013, 2014) proposes a Bartlett type correction for the gradient test, Lemonte (2013) developed the formulae of the gradient test for generalized linear models with dispersion covariates, Lemonte et al. (2012)

studied local power of the gradient test in comparison to the likelihood, Wald, Rao tests, Ferrari & Pinheiro (2014) evaluated the small-sample properties of the gradient test for extreme-value regression models and Medeiros et al. (2014) studied the performance of the gradient test for accelerated failure time models.

The random effect is a statistical concept conceived to accommodate an eventual extra variability due to unknown causes, such as omitted or unobserved variables, measurement error or model misspecification. Models with random effects represent a flexible class through which overdispersion and variance component models can be considered due to the special dependency structure in the variables. Given the *stochastic* nature of the random effects, we have to make assumptions concerning its distribution for inferential purposes. Notation-wise, let $y$ be our sample and the marginal likelihood $m(y|\theta)$ of the model with random effects represented by

$$m(y|\theta) = \int f(y|\theta, z)g(z)dz$$

where $f(y|\theta, z)$ is the conditional likelihood for the parameter $\theta$ which depends on the random effect $z$ with unknown density $g(z)$. This $m(y|\theta)$ is the likelihood of a *mixture model* (Aitkin, 1996a). The assumption of normally distributed random effects is appropriate for many applications, but this also implies that the integration problem of $m(y|\theta)$ is analytically solvable only for conjugated distributions. Numerical methods, such as *Gaussian quadrature* (Golub & Welsch, 1969), are often used or the likelihood function is indirectly maximized.

The main issue of assuming any parametric distribution for the random effects is that this appears very artificial and is difficult to motivate in practice. If it is not possible to make concrete assumptions about the distribution of the random effects, it would be useful to estimate the parameters alongside the density $g(\ \cdot\ )$. A reference of this approach can be found in Anderson & Hinde (1988), where the iterative EM algorithm of Dempster et al. (1977) is used as an indirect method for normally distributed mixtures of Poisson variable. Aitkin & Francis (1995) offer GLIM macros, which calculate the estimators for response distributions from exponential family with unspecified distribution. Application of this technique for the

analysis of overdispersion in generalised linear models (GLMs) is given in Anderson (1988) and Aitkin (1994, 1996a).

In this sense, assuming that $g(\ \cdot\ )$ is *unspecified*, Laird (1978) proposed an estimation method called *Nonparametric Maximum Likelihood* (NPML) which consists in estimating $z$ and $g(z)$ alongside to $\theta$ using an EM algorithm (Hinde, 1982).

Our main goal in this thesis is to evaluate the gradient test on the context of the generalised linear models with random effects. We developed the unified formulae for the gradient test for the models with random effects with normal and unspecified distribution. We performed an extensive Monte Carlo simulation experiment for verifying the type I and power of the gradient test for finite samples. We also present numerical applications to real data sets.

## 1.1 Organisation of the Thesis

In organizing this thesis, we have divided the work in four main chapters. The Chapter 2 establishes the background to this work, giving a comprehensive overview of the asymptotic theory for the likelihood based inference methods and tests. We also express the general definition of the classic asymptotic tests and the gradient test.

In Chapter 3 we define the gradient statistic for testing parameters related to the fixed effects part of the model. For this we define, based on the literature, the generalised linear model with random effects. We also propose a compact matrix notation not seen in the literature before. This notation helps in the development of the R code use latter for simulation and application purposes.

In Chapter 4 we propose the formulae for the Fisher information for generalised linear models with random effects. The proposed formulae includes an analytic method for the model with Gaussian random effects. We also propose an alternative method based on the last EM algorithm estimates which can be applied for either models with Gaussian or unspecified distribution for the random effects. We provide simulation results and an illustrative example. Although the gradient statistic does not use the Fisher information in its formula, we have developed it to obtain the

Wald and Rao statistics for comparison purposes in the Chapter 5.
In Chapter 5 we present an extensive simulation experiment to verify the finite sample properties of the gradient test and compare to the likelihood ratio, Wald and Rao tests. This simulation covered various scenarios of the generalised linear models with random effects including different sample sizes, response distributions, number of mass points and random effects distribution. We also provide four illustrative real data examples for the gradient test.
The Chapter 6 concludes the thesis presenting an overview and discussing the findings of this work.
The Appendix A present the functions in R code used to compute the tests and B gives the analytic or approximated formulae to estimate the variance under the model with normal random effects.

## 1.2 Spin-off publications

Partial results of this thesis have been presented and published in the following conference proceedings.

- da Silva-Júnior, A. H. M., Einbeck, J. & Craig, P. S. (2015). The gradient test for generalised linear models with random effects. In A. Blanco-Fernandez & G. Gonzalez-Rodriguez, eds., *Programme and Abstracts: 9th International Conference on Computational and Financial Econometrics (CFE 2015) and 8th International Conference of the ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on Computing & Statistics (ERCIM 2015)*. 63

- da Silva-Júnior, A. H. M., Einbeck, J. & Craig, P. S. (2016). Gradient test on generalised linear models with random effects. In J.-F. Dupuy & J. Josse, eds., *Proceedings of the 31st International Workshop on Statistical Modelling*, vol. 1. 213–218

- da Silva-Júnior, A. H. M. (2017). Gradient test for variance component models. In M. Grzegorczyk & G. Ceoldo, eds., *Proceedings of the 32nd Inter-*

*national Workshop on Statistical Modelling*, vol. 2. 71–74

Chapter 4 is the outcome of a research project done together with Dr. Jochen Einbeck and Prof. Peter Craig accepted for publication in

da Silva-Júnior, A. H. M., Einbeck, J. & Craig, P. S. (2017). Fisher information on Gaussian quadrature models. *Statistica Neerlandica* In press.

# Chapter 2

# Basics of likelihood inference and the gradient test

## 2.1 Introduction

The method of maximum likelihood has been used extensively to estimate parameters in a large variety of models. The likelihood theory lends properties that allow the formulation of asymptotic hypothesis testing, for instance, likelihood ratio (LR), developed by Wilks et al. (1938), followed by the Wald, (Wald, 1943) and Rao test (Rao, 1948). Such tests have in common the $\chi^2$ as reference distribution for the sample size $n \to \infty$ and under the null hypothesis.

Recently, a new statistic was proposed by Terrell (2002) and has been called *gradient statistic* or *Terrell test*. The gradient statistic is rather simple to compute and does not involve any matrix computations such as matrix products or inversions. The gradient statistic shares the same asymptotic properties of first order with the three previous statistics. These features make the gradient statistic able to compete with the three well-established classical asymptotic tests.

We will suppose, initially, the following situation for the construction of the hypotheses test. Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ a sample of $n$ independent observations of a random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$, which has the pdf $f(\,\cdot\,; \boldsymbol{\theta})$ indexed by an unknown $p$-dimensional vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$. The likelihood function

corresponding to the observed vector $\boldsymbol{y}$ from the density $f(\boldsymbol{y}, \boldsymbol{\theta})$ is written

$$\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{y}) \equiv \mathrm{L}(\boldsymbol{\theta}) = f(\boldsymbol{y}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i, \boldsymbol{\theta}).$$

and the log-likelihood function becomes

$$\ell(\boldsymbol{y}, \boldsymbol{\theta}) \equiv \ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(y_i; \boldsymbol{\theta}). \tag{2.1.1}$$

From (2.1.1) comes the score vector, the (observed) information matrix and the Fisher information matrix defined, respectively, as

$$\begin{aligned}
&\mathbf{U}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}, \quad \mathbf{J}(\boldsymbol{\theta}) = -\frac{\partial \mathbf{U}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}, \quad \text{and} \\
&\mathbf{K}(\boldsymbol{\theta}) = \mathrm{E}\left[\mathbf{U}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})^\top\right] = -\mathrm{E}\left[\frac{\partial \mathbf{U}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right] = \mathrm{E}[\mathbf{J}(\boldsymbol{\theta})].
\end{aligned}$$

The *maximum likelihood estimator* (MLE) is defined as the unique solution to

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{y}; \boldsymbol{\theta}),$$

where $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{y})$ and, if it exists, usually can be obtained by solving the equation $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$ also known as *likelihood equations*. In effect, the sufficient conditions to the existence and uniqueness of a MLE depend on the nature of both $\Theta$ and $\ell(\boldsymbol{\theta})$. If $\Theta$ is a compact space and $\ell(\boldsymbol{\theta})$ is continuous in $\Theta$ then there exists a MLE. Also, if the MLE exists, it is unique when $\Theta$ is a convex space and if $\ell(\,\cdot\,, \boldsymbol{\theta})$ is strictly concave in $\boldsymbol{\theta}$.

Important inferential tools for the MLE are obtained via Taylor series expansion of $\ell(\boldsymbol{\theta})$ and $\mathbf{U}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$. In this sense, there are conditions to be verified in order to discuss the asymptotic properties of the MLE and its functions. Such conditions are often called *regularity conditions* and will be presented in detail in Section 2.3.

## 2.2 Basic concepts of convergence

First, let $\{Y_n\}$ be a sequence of random variables defined for a large $n$. Here $n$ does not necessarily represent the sample size. We then present some important stochastic convergences that will be used on the next sections.

### 2.2.1 Convergence in probability

The sequence $\{Y_n\}$ *converges in probability* for a random variable $Y$ (which can be degenerate) if

$$\lim_{n\to\infty} \Pr(|Y_n - Y| < \epsilon) = 1$$

for all $\epsilon > 0$. This convergence is denoted by $Y_n \xrightarrow{\mathcal{P}} Y$ and means that $Y_n$ and $Y$ are approximately equal with probability close to 1 for a sufficiently large $n$.

### 2.2.2 Almost sure convergence

The sequence $\{Y_n\}$ converges *almost surely* to a random variable $Y$ if

$$\Pr\left(\lim_{n\to\infty} Y_n = Y\right) = 1.$$

We denote this convergence by $Y_n \xrightarrow{\text{a.s.}} Y$.

### 2.2.3 Convergence in distribution

The sequence $\{Y_n\}$ *converges in distribution* to $Y$ if

$$\lim_{n\to\infty} \Pr(Y_n < y) = F_Y(y),$$

for every $y$ in $\mathbb{R}$ where the distribution function $F_Y(\,\cdot\,)$ of $Y$ is continuous. We denote this convergence by $Y_n \xrightarrow{\mathcal{D}} Y$.

### 2.2.4 Mann-Wald notation

The Mann-Wald notation is useful for describing the order of magnitude of specified quantities.

Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of positive values and $\{Y_n\}_{n=1}^{\infty}$ a sequence of random vectors. We denote

$Y_n = \mathcal{o}_p(a_n)$ which means that $a_n^{-1} Y_n \xrightarrow{\mathscr{P}} \mathbf{0}_p$, where $\mathbf{0}_p$ is a vector in $\mathbb{R}^p$, and

$Y_n = \mathcal{O}_p(a_n)$ which means that, for any $\epsilon > 0$ there exist $\kappa < \infty$ and $n_0 < \infty$ such that, for all $n > n_0$

$$\Pr[\|a_n^{-1} Y_n\| > \kappa] < \epsilon.$$

## 2.3 Regularity conditions

The following regularity conditions are used in asymptotic theory to justify and define the error terms of Taylor series expansions. Some of these conditions or all of them are necessary to prove the asymptotic properties of the MLE such as consistency, normality and efficiency.

First, assume that $\boldsymbol{y}$ is a realisation of a random vector $\mathbf{Y}$ with distribution $P_{\boldsymbol{\theta}}$ which belongs to a class $\mathscr{P}$ and depends on $\boldsymbol{\theta} \in \Theta$. Also, the observations $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, where $y_i$ are *iid* with density $f(y_i, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

The following assumptions will be required further in this chapter:

(i) the distributions $P_{\boldsymbol{\theta}}$ are distinct, i.e., $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$ implies $P_{\boldsymbol{\theta}} \neq P_{\boldsymbol{\theta}'}$;

(ii) the distributions $f(\,\cdot\,, \boldsymbol{\theta})$ have common support for all $\boldsymbol{\theta} \in \Theta$, i.e., the set $\mathscr{A}_{\boldsymbol{\theta}} = \{\boldsymbol{y}; f(\boldsymbol{y}, \boldsymbol{\theta}) > 0\}$ does not depend on $\boldsymbol{\theta}$;

The condition (i) ensures that the probability distributions are different for distinct $\boldsymbol{\theta}$ and for the given data. The condition (ii) ensures that the sample space of $\boldsymbol{y}$ is identical and is independent of $\boldsymbol{\theta}$.

Consider the observations $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, where $y_i$ are *iid* with density $f(y_i, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

The assumptions (iii) and (iv) below ensure the regularity of $f(\boldsymbol{y}, \boldsymbol{\theta})$ as function of $\boldsymbol{\theta}$ and the existence of an open set $\Theta_1$ in the parametric space $\Theta$ such as the true parameter $\theta_0$ belongs to $\Theta_1$:

(iii) there exists an open set $\Theta_1 \subset \Theta$ which contains $\boldsymbol{\theta}_0$ such that the density function $f(\boldsymbol{y}, \boldsymbol{\theta})$, for almost all $\boldsymbol{y}$, which admits all the derivatives until third order in relation to $\boldsymbol{\theta}$, for all $\boldsymbol{\theta} \in \Theta_1$;

(iv) $\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{U}(\boldsymbol{\theta})] = 0$ and the information matrix $\mathbf{K}(\boldsymbol{\theta})$ is positive definite and has finite values for all $\boldsymbol{\theta} \in \Theta_1$;

(v) there are functions $\mathscr{M}_{ijk}(\boldsymbol{y})$ which shall not depend on $\boldsymbol{\theta}$ such that, for $i$, $j$ and $k = 1, \ldots, p$,

$$\left| \frac{\partial^3 f(\boldsymbol{y}; \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < \mathscr{M}_{ijk}(\boldsymbol{y})$$

for all $\boldsymbol{\theta} \in \Theta_1$, where $\mathrm{E}_{\boldsymbol{\theta}_0}[\mathscr{M}_{ijk}(Y)] < \infty$.

The condition (iii) represents the existence of $\Theta_1$ and the derivatives of $f(\boldsymbol{y}; \boldsymbol{\theta})$ until third order in $\Theta_1$. The condition (iv) ensures that the information matrix is finite and positive-definite in an open neighbourhood of $\boldsymbol{\theta}_0$. Finally, the condition (v) ensures that the third order derivatives of the log-likelihood are bounded by a integrable function of $\mathbf{Y}$ whose expected value is finite (Cordeiro, 1999).

The models discussed on Chapter 3 make use of the *mixture models* theory from Aitkin (1996a) for modelling the random effects. In this sense, we have some considerations about the regularity conditions stated above. According to Chen & Li (2009), the regularity conditions (i), (iv) and (v) are not always valid for *Gaussian mixture models*. We have then some undesired consequences such as *unbounded likelihood function*, *loss of strong identifiability* and *infinite Fisher information*. However, this might not be an issue as we do not intent to test parameters regarding to the random effects.

On the other hand, for *non-parametric maximum likelihood mixture models*, Lindsay (1995, chap. 1, pg. 24) makes the remark: "one of the most striking features of the above theory [Nonparametric maximum likelihood estimation] is the complete lack of regularity conditions on the models and the complete generality with regard to the parameter space of $\phi$".

## 2.4 Asymptotic properties of the MLE

### 2.4.1 Consistency

Commonly an estimator is considered a function of the sample size $n$ and, as long as we increase it ($n \to \infty$), we intuitively expect an enhance of the estimator precision.

**Definition 2.4.1** Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ be *iid* with density $f(y_i, \boldsymbol{\theta})$ for each $y_i$, $i = 1, \ldots, n$. Then, for $n \to \infty$, an estimate $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(\boldsymbol{y})$ is considered *consistent* for the parameter $\boldsymbol{\theta}$ if it satisfies

$$\lim_{n\to\infty} \mathrm{MSE}(\hat{\boldsymbol{\theta}}_n) = \mathbf{0},$$

where $\mathrm{MSE}(\hat{\boldsymbol{\theta}}_n) = \mathrm{E}[(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^\top(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})]$ is the *mean square error* of $\hat{\boldsymbol{\theta}}_n$.

In general, two definitions of consistency are widely used in asymptotic theory.

**Definition 2.4.2** *weak consistency*: Let $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(\boldsymbol{y})$ the estimator for $\boldsymbol{\theta}$ based on the iid sample $\boldsymbol{y}$. Then, $\hat{\boldsymbol{\theta}}_n$ is *weakly consistent* if, for $n \to \infty$

$$\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + \mathcal{o}_p(1)\,.$$

**Definition 2.4.3** *strong consistency*: Let $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(\boldsymbol{y})$ the estimator for $\boldsymbol{\theta}$ based on the iid sample $\boldsymbol{y}$. Then, $\hat{\boldsymbol{\theta}}_n$ is *strongly consistent* if, for $n \to \infty$

$$\Pr\left[\lim_{n\to\infty} \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| = 0\right] = 1.$$

This means that the weak or strong consistency happens when $\hat{\boldsymbol{\theta}}_n$ satisfies the weak law or the strong law of large numbers, respectively.

### 2.4.2 Normality

**Theorem 2.4.4** Assume the *iid* sample $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ with density $f(y_i, \boldsymbol{\theta})$ and regularity conditions (i)–(v) valid. If $\tilde{\boldsymbol{\theta}}$ is a consistent solution for the maximum

likelihood equations $\mathbf{U}(\boldsymbol{\theta}) = 0$, then

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \mathbf{k}(\boldsymbol{\theta}_0)^{-1}). \tag{2.4.2}$$

In other words, for large sample sizes, the distribution of $\tilde{\boldsymbol{\theta}}$ is approximately $p$-dimensional normal with mean $\boldsymbol{\theta}_0$ and covariance matrix $\mathbf{K}(\boldsymbol{\theta}_0)^{-1} = n^{-1}\mathbf{K}(\boldsymbol{\theta}_0)^{-1}$. Cramér (1999, Sec 33.3) and Lehmann & Casella (1998, Sec 6.4) show rigorous demonstrations of convergence of (2.4.2) for $p = 1$ and $p \geqslant 1$, respectively.

We shall demonstrate (2.4.2) for the uniparametric case. The general regularity conditions ensure the expansion of $\mathbf{U}(\tilde{\theta}) = 0$ around the true parameter $\boldsymbol{\theta}_0$ up to second order

$$\mathbf{U}(\theta_0) + \mathbf{U}'(\theta_0)(\tilde{\theta} - \theta_0) + \frac{1}{2}\mathbf{U}''(\theta^*)(\tilde{\theta} - \theta_0)^2 = 0$$

where $|\theta^* - \theta_0| < |\tilde{\theta} - \theta_0|$ and, therefore, $\theta^*$ is necessarily consistent for $\theta_0$. The first two terms on the left side of the equation are $\mathcal{O}_p\left(n^{1/2}\right)$ and the third is $\mathcal{O}_p(1)$, as $\mathbf{U}'(\theta_0) = \mathcal{O}_p(n)$, $\mathbf{U}''(\theta_0) = \mathcal{O}_p(n)$ and $\tilde{\theta} - \theta_0 = \mathcal{O}_p\left(n^{-1/2}\right)$. The $\mathbf{U}(\theta_0)$ and $\mathbf{U}'(\theta_0)$ are sums of *iid* random variables so the expansion implies

$$\sqrt{n}(\hat{\theta} - \theta_0)\left\{-\sum_{i=1}^{n}\frac{\mathbf{U}_i(\theta_0)}{n\mathbf{K}(\theta_0)} + \mathcal{O}_p\left(n^{-1/2}\right)\right\} = \sum_{i=1}^{n}\frac{\mathbf{U}_i(\theta_0)}{\sqrt{n}\mathbf{K}(\theta_0)}.$$

where $\mathbf{K}(\theta_0) = n^{-1}\mathbf{K}(\theta_0)$ is the information of a single observation. By the weak law of large numbers, $\sum_{i=1}^{n} n^{-1}\mathbf{U}'(\theta_0)/\mathbf{K}(\theta_0) = 1 + \mathcal{o}_p(1)$. Then,

$$\sqrt{n}(\hat{\theta} - \theta_0)\{1 + \mathcal{o}_p(1)\} = \sum_{i=1}^{n}\frac{\mathbf{U}_i(\theta_0)}{\sqrt{n}\mathbf{K}(\theta_0)}. \tag{2.4.3}$$

From (2.4.3) and the condition (iv), we can prove that $\hat{\theta}$ has asymptotic mean equal 0 and covariance structure given by $\mathbf{K}(\theta_0)$. Thus, the asymptotic normality of $\hat{\theta}$ is obtained via central limit theorem applied to the right side of (2.4.3).

### 2.4.3 Efficiency

An estimator $\hat{\boldsymbol{\theta}}$ is considered *asymptotically efficient* for $\boldsymbol{\theta}$ if it is consistent, asymptotically normal and its covariance matrix is no larger than the covariance matrix

of any other estimator $\boldsymbol{\theta}^* \in \Theta$ which is consistent and asymptotically normal.
The results of asymptotic efficiency and asymptotic normality can be generalised for less restritive cases such as mixture models, provided that by week law of large numbers $n^{-1}\mathbf{J}(\boldsymbol{\theta}) \xrightarrow{\mathscr{P}} n^{-1}\mathbf{K}(\boldsymbol{\theta})$ (Liang, 1984; Lindsay et al., 1991; Bickel et al., 1993).

## 2.5 The gradient test and the classical asymptotic tests

We present here the general idea of the gradient test proposed by Terrell (2002) and its older sister tests, the likelihood ratio, Wald and Rao.

### 2.5.1 Simple hypothesis

Our chief concern will be testing the null hypothesis $\mathscr{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative hypothesis $\mathscr{H}_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is an arbitrary vector.
The definitions of the likelihood ratio, Wald and score test statistics for $\mathscr{H}_0$ are, respectively,

$$\begin{aligned}
\xi_{\mathscr{LR}} &= 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0)], \\
\xi_{\mathscr{W}} &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{K}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \\
\xi_{\mathscr{R}} &= \mathbf{U}(\boldsymbol{\theta}_0)^\top \mathbf{K}(\boldsymbol{\theta}_0)^{-1} \mathbf{U}(\boldsymbol{\theta}_0),
\end{aligned}$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$, which can be obtained by $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$. A different approach for the Wald test is to substitute the Fisher information matrix estimated under the alternative hypothesis by the theoretical equivalent under null hypothesis. Here, we will call this approach as modified Wald statistic, and define as

$$\xi_{\mathscr{MW}} = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{K}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

Cordeiro (1999) shows that the asymptotic distributions of $\xi_{\mathscr{W}}$, $\xi_{\mathscr{R}}$ and $\xi_{\mathscr{MW}}$ can be

obtained considering that

$$\begin{aligned}\sqrt{n}\mathbf{U}(\boldsymbol{\theta}) &\xrightarrow{\mathscr{D}} \mathcal{N}_p(\mathbf{0}, \bar{\mathbf{K}}(\boldsymbol{\theta}))\\ \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &\xrightarrow{\mathscr{D}} \mathcal{N}_p(\mathbf{0}, \bar{\mathbf{K}}(\boldsymbol{\theta})^{-1}).\end{aligned}$$

where $\mathbf{K}(\boldsymbol{\theta}) = n\bar{\mathbf{K}}(\boldsymbol{\theta})$. If $\mathbf{K}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$ thus, for $n \to \infty$,

$$\begin{aligned}n^{-1}\mathbf{J}(\boldsymbol{\theta}^{(0)}) &\xrightarrow{\mathscr{P}} \bar{\mathbf{K}}(\boldsymbol{\theta}^{(0)})\\ n^{-1}\mathbf{J}(\hat{\boldsymbol{\theta}}) &\xrightarrow{\mathscr{P}} \bar{\mathbf{K}}(\boldsymbol{\theta}^{(0)}).\end{aligned} \tag{2.5.4}$$

One can show that $\xi_{\mathscr{LR}}$ has chi-squared distribution using the Taylor expansion of $\ell(\boldsymbol{\theta}^{(0)})$ around the solution $\hat{\boldsymbol{\theta}}$ from $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ and (2.5.4). Thus,

$$\begin{aligned}\ell(\boldsymbol{\theta}^{(0)}) &= \ell(\hat{\boldsymbol{\theta}}) + \cancelto{0}{\mathbf{U}(\hat{\boldsymbol{\theta}})}(\boldsymbol{\theta}^{(0)} - \hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta}^{(0)} - \hat{\boldsymbol{\theta}})^\top \mathbf{J}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \mathcal{O}_p(1)\,,\\ &= \ell(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta}^{(0)} - \hat{\boldsymbol{\theta}})^\top \mathbf{K}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(0)} - \hat{\boldsymbol{\theta}}) + \mathcal{O}_p(1)\end{aligned}$$

or

$$\xi_{\mathscr{LR}} = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})^\top \mathbf{K}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}) + \mathcal{O}_p(1)\,. \tag{2.5.5}$$

Likewise, the Taylor expansion for $\hat{\boldsymbol{\theta}}$ around $\boldsymbol{\theta}^{(0)}$

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \boldsymbol{\theta}^{(0)} + \mathbf{K}(\boldsymbol{\theta}^{(0)})^{-1}\mathbf{U}(\boldsymbol{\theta}^{(0)}) + \mathcal{O}_p\left(n^{-1/2}\right)\\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} &= \mathbf{K}(\boldsymbol{\theta}^{(0)})^{-1}\mathbf{U}(\boldsymbol{\theta}^{(0)}) + \mathcal{O}_p\left(n^{-1/2}\right)\end{aligned} \tag{2.5.6}$$

Substituting (2.5.6) in (2.5.5), we have

$$\begin{aligned}\xi_{\mathscr{LR}} &= [\mathbf{K}(\boldsymbol{\theta}^{(0)})^{-1}\mathbf{U}(\boldsymbol{\theta}^{(0)})]^\top \mathbf{K}(\hat{\boldsymbol{\theta}})[\mathbf{K}(\boldsymbol{\theta}^{(0)})^{-1}\mathbf{U}(\boldsymbol{\theta}^{(0)})] + \mathcal{O}_p(1)\\ &= \mathbf{U}(\boldsymbol{\theta}^{(0)})^\top [\mathbf{K}(\boldsymbol{\theta}^{(0)})^{-1}]^\top \mathbf{K}(\hat{\boldsymbol{\theta}})\mathbf{K}(\boldsymbol{\theta}^{(0)})^{-1}\mathbf{U}(\boldsymbol{\theta}^{(0)}) + \mathcal{O}_p(1)\,,\end{aligned}$$

where commonly $\mathbf{K}(\,\cdot\,)$ is a symmetric matrix, then $\mathbf{K}(\,\cdot\,)^\top = \mathbf{K}(\,\cdot\,)$ (also valid for its inverse) and by the strong consistency of $\hat{\boldsymbol{\theta}}$, $\mathbf{K}(\hat{\boldsymbol{\theta}}) \xrightarrow{\text{a.s.}} \mathbf{K}(\boldsymbol{\theta}^{(0)})$, thus

$$\xi_{\mathscr{LR}} = \mathbf{U}(\boldsymbol{\theta}^{(0)})^\top \mathbf{K}(\boldsymbol{\theta}^{(0)})^{-1}\mathbf{U}(\boldsymbol{\theta}^{(0)}) + \mathcal{O}_p(1)\,.$$

The statistics $\xi_{\mathscr{LR}}$, $\xi_{\mathscr{W}}$, $\xi_{\mathscr{R}}$ and $\xi_{\mathscr{MW}}$ have centred chi-square distribution approximately with $p$ degrees of freedom ($\chi^2_p$) under the null hypothesis $\mathscr{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. Therefore, we reject $\mathscr{H}_0$ if the observed value of the statistic exceeds the quantile $100 \times (1-\alpha)\%$ of the $\chi^2_p$ distribution, with nominal level $\alpha$.

We are now able to discuss the idea behind the gradient statistic. Let $\mathbf{M}_{p\times p}$ a square matrix that satisfies the condition $\mathbf{M}^\top\mathbf{M} = \mathbf{K}(\boldsymbol{\theta})$. Using this matrix, we can rewrite $\xi_{\mathscr{R}}$ and $\xi_{\mathscr{MW}}$ as

$$\begin{aligned}\xi_{\mathscr{R}} &= [(\mathbf{M}^{-1})^\top\mathbf{U}(\boldsymbol{\theta}_0)]^\top(\mathbf{M}^{-1})^\top\mathbf{U}(\boldsymbol{\theta}_0),\\ \xi_{\mathscr{MW}} &= [(\mathbf{M})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)]^\top\mathbf{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0).\end{aligned}$$

Lemonte (2016) shows that

$$\begin{aligned}(\mathbf{M}^{-1})^\top\mathbf{U}(\boldsymbol{\theta}_0) &\sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p),\\ \mathbf{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) &\sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p),\end{aligned}$$

where $\mathbf{I}_p$ is a $p$-dimensional identity matrix.

Furthermore, the inner product between $(\mathbf{M}^{-1})^\top\mathbf{U}(\boldsymbol{\theta}_0)$ and $\mathbf{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ results in

$$\begin{aligned}[(\mathbf{M}^{-1})^\top\mathbf{U}(\boldsymbol{\theta}_0)]^\top\mathbf{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) &= \mathbf{U}(\boldsymbol{\theta}_0)^\top\mathbf{M}^{-1}\mathbf{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\\ &= \mathbf{U}(\boldsymbol{\theta}_0)^\top(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0).\end{aligned}$$

Based on the last expression, we have the following definition:

**Definition 2.5.1** (Terrell, 2002) The gradient statistic, $\xi_{\mathscr{T}}$, to test the simple null hypothesis $\mathscr{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $\mathscr{H}_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ has the form

$$\xi_{\mathscr{T}} = \mathbf{U}(\boldsymbol{\theta}_0)^\top(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0).$$

**Theorem 2.5.2** Under $\mathscr{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\xi_{\mathscr{T}}$ has $\chi^2_p + o_p(1)$ distribution.

**Proof**: The MLE $\hat{\boldsymbol{\theta}}$ is asymptotically efficient under the regularity conditions and

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathbf{K}(\boldsymbol{\theta}_0)^{-1}\mathbf{U}(\boldsymbol{\theta}_0) + o_p\left(n^{-1/2}\right).$$

We already know that, in same conditions,

$$\mathbf{U}(\boldsymbol{\theta}_0) = \mathcal{O}_p\left(n^{1/2}\right),$$

then

$$\xi_{\mathcal{T}} = \mathbf{U}(\boldsymbol{\theta}_0)^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathbf{U}(\boldsymbol{\theta}_0)\mathbf{K}(\boldsymbol{\theta}_0)^{-1}\mathbf{U}(\boldsymbol{\theta}_0) + \mathcal{O}_p(1) = \xi_{\mathcal{R}}.$$

Therefore, as $\xi_{\mathcal{R}}$ has $\chi^2_p + \mathcal{O}_p(1)$ then $\xi_{\mathcal{T}}$ has as well. □

Note that $\xi_{\mathcal{T}}$ has the advantage of not involving the estimated Fisher information matrix neither its inverse. We cannot state that $\xi_{\mathcal{T}}$ is non-negative for any scenario except for the case stated in the Theorem 2.5.3.

**Theorem 2.5.3** (Terrell, 2002) If $\ell(\boldsymbol{\theta})$ is uni-modal and differentiable in $\boldsymbol{\theta}$, so

$$\xi_{\mathcal{T}} = \mathbf{U}(\boldsymbol{\theta}_0)^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \geqslant 0$$

**Proof**: Assuming the regularity conditions (i)–(v) and by the uniqueness of the MLE, $\hat{\boldsymbol{\theta}}$ is the only existent point of maxima of $\ell(\,\cdot\,)$ and therefore, solution for $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$. Let exist a $\boldsymbol{\theta}' = (\theta'_1, \ldots, \theta'_p)^\top \in \Theta$ such that

$$\mathbf{U}(\boldsymbol{\theta}')^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}') < 0,$$

i.e., a violation of the Theorem. Then, $\mathbf{U}(\boldsymbol{\theta}') \neq \mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ and $\boldsymbol{\theta}' \neq \hat{\boldsymbol{\theta}}$. This means that only $\boldsymbol{\theta}' < \hat{\boldsymbol{\theta}}$ or $\boldsymbol{\theta}' > \hat{\boldsymbol{\theta}}$ might be true. If $\theta'_i < \hat{\theta}_i$, for $i$ in $1, \ldots, p$, then $\mathbf{U}(\theta'_i) < \mathbf{U}(\hat{\theta}_i)$ by the uni-modality of $\ell(\,\cdot\,)$. As a consequence,

$$\mathbf{U}(\boldsymbol{\theta}')^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}') > 0.$$

For the second situation, if $\theta'_i > \hat{\theta}_i$, for $i$ in $1, \ldots, p$, then $\mathbf{U}(\theta'_i) > \mathbf{U}(\hat{\theta}_i)$ also by the uni-modality of $\ell(\,\cdot\,)$. Thus,

$$\mathbf{U}(\boldsymbol{\theta}')^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}') > 0.$$

Therefore, we prove by contradiction that $\xi_{\mathcal{T}} \geqslant 0$ if $\ell(\boldsymbol{\theta})$ is uni-modal. □

**Example 2.5.1** Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ a $n$ size random sample from a Gaussian distribution with mean $\theta$ and variance 1, $\mathcal{N}(\theta, 1)$. Thus,

$$\begin{aligned}\ell(\theta) &= \log\left[\prod_{i=1}^{n}(2\pi)^{-1/2}\exp\left\{-\frac{(y_i-\theta)^2}{2}\right\}\right]\\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2,\end{aligned}$$

which provides the uni-parametric versions of score and Fisher information, respectively

$$\mathbf{U}(\theta) = \sum_{i=1}^{n} y_i - n\theta, \quad \mathbf{K}(\theta) = n.$$

so that for $\mathbf{U}(\theta) = 0$, the maximum likelihood estimator for $\theta$ is $\hat{\theta} = \sum_{i=1}^{n} y_i/n = \bar{y}$. Consider the null hypothesis $\mathscr{H}_0 : \theta = \theta_0$. For testing $\mathscr{H}_0$, the likelihood ratio statistic assumes

$$\begin{aligned}\xi_{\mathscr{LR}} &= 2\left[-\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_0)^2\right]\\ &= \sum_{i=1}^{n}[(y_i-\theta_0)^2 - (y_i-\bar{y})^2] = \sum_{i=1}^{n}[y_i^2 - 2\theta_0 y_i + \theta_0^2 - y_i^2 + 2\bar{y}y_i - \bar{y}^2]\\ &= -2n\theta_0\bar{y} + n\theta_0^2 + 2n\bar{y}^2 - n\bar{y}^2 = n[\bar{y}^2 - 2\theta_0\bar{y} + \theta_0^2]\\ &= n(\bar{y}-\theta_0)^2.\end{aligned}$$

Similarly, the Wald, Rao and gradient statistics are, respectively,

$$\begin{aligned}\xi_{\mathscr{W}} &= (\hat{\theta}-\theta_0)^2\mathbf{K}(\hat{\theta})\\ &= n(\bar{y}-\theta_0)^2,\end{aligned}$$

$$\begin{aligned}\xi_{\mathscr{R}} &= \mathbf{U}(\theta_0)^2/\mathbf{K}(\theta_0)\\ &= [n(\bar{y}-\theta_0)]^2/n\\ &= n(\bar{y}-\theta_0)^2,\end{aligned}$$

and

$$\begin{aligned}\xi_{\mathscr{T}} &= \mathbf{U}(\theta_0)(\hat{\theta}-\theta_0)\\ &= n(\bar{y}-\theta_0)(\bar{y}-\theta_0)\\ &= n(\bar{y}-\theta_0)^2.\end{aligned}$$

**Example 2.5.2** Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ a $n$ size random sample from a exponential distribution with pdf

$$f(y;\theta) = \frac{1}{\theta}\exp\left\{-\frac{y}{\theta}\right\}.$$

Thus,

$$\begin{aligned}\ell(\theta) &= \log\left[\prod_{i=1}^{n}\frac{1}{\theta}\exp\left\{-\frac{y_i}{\theta}\right\}\right]\\ &= -n\log\theta - \frac{1}{\theta}\sum_{i=1}^{n} y_i.\end{aligned}$$

which provides the respectively uni-parametric versions of score and Fisher information

$$\mathbf{U}(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} y_i, \quad \mathbf{K}(\theta) = \frac{n}{\theta^2}.$$

so that for $\mathbf{U}(\theta) = 0$, the maximum likelihood estimator for $\theta$ is $\hat{\theta} = \sum_{i=1}^{n} y_i/n = \bar{y}$. Consider the null hypothesis $\mathscr{H}_0 : \theta = \theta_0$. For testing $\mathscr{H}_0$, the likelihood ratio statistic is

$$\begin{aligned}\xi_{\mathscr{LR}} &= 2\left[-n\log\bar{y} - \frac{1}{\bar{y}}\sum_{i=1}^{n} y_i + n\log\theta_0 + \frac{1}{\theta_0}\sum_{i=1}^{n} y_i\right]\\ &= 2n\left[\log\left(\frac{\theta_0}{\bar{y}}\right) + \frac{\bar{y}}{\theta_0} - 1\right].\end{aligned}$$

Similarly, the Wald, Rao and gradient statistics are, respectively,

$$\begin{aligned}\xi_{\mathscr{W}} &= (\hat{\theta}-\theta_0)^2 k(\hat{\theta})\\ &= n(\bar{y}-\theta_0)^2\frac{n}{\bar{y}^2}\\ &= n\left(\frac{\bar{y}-\theta_0}{\bar{y}}\right)^2\end{aligned}$$

$$\begin{aligned}\xi_{\mathscr{R}} &= \mathbf{U}(\theta_0)^2/\mathbf{K}(\theta_0) \\ &= \left[\frac{n}{\theta_0}\left(\frac{\bar{y}}{\theta_0}-1\right)\right]^2 \Big/ \left(\frac{n}{\theta_0^2}\right) \\ &= n\left(\frac{\bar{y}-\theta_0}{\theta_0}\right)^2,\end{aligned}$$

and

$$\begin{aligned}\xi_{\mathscr{T}} &= \mathbf{U}(\theta_0)(\hat{\theta}-\theta_0) \\ &= n\left(\frac{\bar{y}}{\theta_0^2}-\frac{1}{\theta_0}\right)(\bar{y}-\theta_0) \\ &= n\left(\frac{\bar{y}^2-\bar{y}\theta_0-\bar{y}\theta_0+\theta_0^2}{\theta_0^2}\right) \\ &= n\left(\frac{\bar{y}-\theta_0}{\theta_0}\right)^2.\end{aligned}$$

### 2.5.2 Composite hypothesis

We now will consider the problem of testing the hypotheses

$$\begin{cases} \mathscr{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)} \\ \mathscr{H}_a : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^{(0)} \end{cases},$$

which implies the partitioning $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ where $\boldsymbol{\theta}_1 = (\theta_1, \ldots, \theta_q)^\top$ is a $q$-dimensional parameter of interest, $\boldsymbol{\theta}_2 = (\theta_{q+1}, \ldots, \theta_p)^\top$ is a $(p-q)$-dimensional nuisance parameter and $\boldsymbol{\theta}_1^{(0)}$ is a specified vector. Let $\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ the log-likelihood for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. The unrestricted maximum likelihood estimator is $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^\top, \hat{\boldsymbol{\theta}}_2^\top)^\top$ and the restricted maximum likelihood estimator of $\boldsymbol{\theta}_2$ under $\mathscr{H}_0$ is written $\tilde{\boldsymbol{\theta}}_2$; so, $\tilde{\boldsymbol{\theta}}^\top = (\boldsymbol{\theta}_1^{(0)\top}, \tilde{\boldsymbol{\theta}}_2^\top)$ represents the estimator of the full parameter vector $\boldsymbol{\theta}$ under the null hypothesis. We make use for further formulae the mathematical accents $\sim$ and $\wedge$ to represent the estimators under null and alternative hypothesis, respectively.

The score vector $\mathbf{U}$, the Fisher information matrix $\mathbf{K}$ and the inverted Fisher infor-

mation matrix $\mathbf{K}^{-1}$ are also partitioned according to $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, i.e.

$$\mathbf{U} \equiv \mathbf{U}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{pmatrix},$$
$$\mathbf{K} \equiv \mathbf{K}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{pmatrix}, \quad \text{and}$$
$$\mathbf{K}^{-1} \equiv \mathbf{K}^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{K}^{11} & \mathbf{K}^{12} \\ \mathbf{K}^{21} & \mathbf{K}^{22} \end{pmatrix},$$

Similarly, we can use the same notation for the observed information matrix $\mathbf{J}$ and its inverse $\mathbf{J}^{-1}$. In general, the $\mathbf{U}_1$, $\mathbf{U}_2$, $\mathbf{K}_{11}$, $\mathbf{K}_{12} = \mathbf{K}_{21}^\top$ and $\mathbf{K}_{22}$ depend on both $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$.

The likelihood ratio statistic for $\mathscr{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$ is

$$\xi_{\mathscr{LR}} = 2[\ell(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) - \ell(\boldsymbol{\theta}_1^{(0)}, \tilde{\boldsymbol{\theta}}_2)]. \tag{2.5.7}$$

The inconvenience of (2.5.7) is that $\xi_{\mathscr{LR}}$ requires two maximisations. One can show that $\xi_{\mathscr{LR}} \xrightarrow{\mathscr{D}} \chi_q^2$ according to Wilks et al. (1938).

The Wald statistic is developed on the basis of the asymptotic normality of the MLE $\hat{\boldsymbol{\theta}}_1$. The idea is that the distribution of $\hat{\boldsymbol{\theta}}$ is, asymptotically, a $p$-dimensional normal distribution, where $\mathbf{K}^{-1}$ is the covariance matrix. Thus, under $\mathscr{H}_0$, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_1$ is also normal, however, $q$-dimensional and with mean $\boldsymbol{\theta}_1^{(0)}$ and covariance matrix $\mathbf{K}^{11}$. This means that, $\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)} \xrightarrow{\mathscr{D}} \mathcal{N}_q(0, \mathbf{K}^{11})$. The matrix $\mathbf{K}^{11}$ can be consistently estimated by $\mathbf{K}^{11}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$, $\mathbf{K}^{11}(\boldsymbol{\theta}_1^{(0)}, \tilde{\boldsymbol{\theta}}_2)$, $\mathbf{J}^{11}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ and $\mathbf{J}^{11}(\boldsymbol{\theta}_1^{(0)}, \tilde{\boldsymbol{\theta}}_2)$. If we choose the first option, the Wald statistic can be expressed by

$$\xi_{\mathscr{W}} = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)})^\top \hat{\mathbf{K}}^{11^{-1}} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}) \tag{2.5.8}$$

where $\hat{\mathbf{K}}^{11} = \mathbf{K}^{11}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$. In (2.5.8), $\xi_{\mathscr{W}}$ is a "quadratic form" which corresponds to an inner product of a two vectors that have the same asymptotically normal distribution $\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)} \xrightarrow{\mathscr{D}} \mathcal{N}_q(\mathbf{0}, \mathbf{K}^{11})$ and, therefore, $\xi_{\mathscr{W}} \xrightarrow{\mathscr{D}} \chi_q^2$ under null hypothesis.

The Rao statistic is based on the asymptotic normality for the score function $\mathbf{U}_1 =$

$\mathbf{U}_1(\boldsymbol{\theta}_1^{(0}, \boldsymbol{\theta}_2)$ applied to the vector of parameters under test, i.e.,

$$\mathbf{U}_1 \xrightarrow{\mathscr{D}} \mathcal{N}_q(\mathbf{0}, \mathbf{K}_{11}), \tag{2.5.9}$$

where $\hat{\mathbf{K}}_{11} = \mathbf{K}_{11}(\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2)$ is the asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}_1$. Thus, the Rao statistic is defined by the quadratic form

$$\xi_{\mathscr{R}} = \tilde{\mathbf{U}}_1^\top \tilde{\mathbf{K}}^{11} \tilde{\mathbf{U}}_1, \tag{2.5.10}$$

where $\tilde{\mathbf{U}}_1 = \mathbf{U}_1(\boldsymbol{\theta}_1^{(0)}, \tilde{\boldsymbol{\theta}}_2)$ and $\tilde{\mathbf{K}}^{11} = \mathbf{K}^{11}(\boldsymbol{\theta}_1^{(0)}, \tilde{\boldsymbol{\theta}}_2)$. The Rao statistic advantage is that it depends only on the MLE under null hypothesis. The asymptotic distribution of $\xi_{\mathscr{R}}$, under $\mathscr{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$, comes directly from (2.5.9) which implies $\xi_{\mathscr{R}} \xrightarrow{\mathscr{D}} \chi^2_q$.
The gradient statistic comes from the results of (2.5.8) and (2.5.10). Let $\mathbf{M}$ a square matrix with dimensions $q \times q$, which satisfies the condition $\mathbf{M}^\top \mathbf{M} = \mathbf{K}^{11}$. Consider the $\xi_{\mathscr{W}}$ version which uses $\tilde{\mathbf{K}}^{11} = \mathbf{K}^{11}(\boldsymbol{\theta}_1^{(0)}, \tilde{\boldsymbol{\theta}}_2)$ to estimate $\mathbf{K}^{11}$. We can rewrite both $\xi_{\mathscr{W}}$ and $\xi_{\mathscr{R}}$ in terms of $\mathbf{M}$ as follows

$$\begin{aligned}
\xi_{\mathscr{W}} &= (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)})^\top \tilde{\mathbf{K}}^{11^{-1}} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}) \\
&= (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)})^\top (\mathbf{M}^\top \mathbf{M})^{-1} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}) \\
&= [(\mathbf{M}^{-1})(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)})]^\top (\mathbf{M}^{-1})(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}), \quad \text{and}
\end{aligned} \tag{2.5.11}$$

$$\begin{aligned}
\xi_{\mathscr{R}} &= \tilde{\mathbf{U}}_1^\top \tilde{\mathbf{K}}^{11} \tilde{\mathbf{U}}_1 \\
&= \tilde{\mathbf{U}}_1^\top \mathbf{M}^\top \mathbf{M} \tilde{\mathbf{U}}_1 \\
&= [(\mathbf{M})^\top \tilde{\mathbf{U}}_1]^\top \mathbf{M}^\top \tilde{\mathbf{U}}_1
\end{aligned} \tag{2.5.12}$$

Both (2.5.11) and (2.5.12) are explicit quadratic forms, so that

$$(\mathbf{M}^{-1})(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}) \xrightarrow{\mathscr{D}} \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q) \tag{2.5.13}$$

$$\mathbf{M}^\top \tilde{\mathbf{U}}_1 \xrightarrow{\mathscr{D}} \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q), \tag{2.5.14}$$

where $\mathbf{I}_q$ is an $q$-dimensional identity matrix. Therefore, the gradient statistic is

result of the inner product between (2.5.13) and (2.5.14), i.e.

$$\begin{aligned}\xi_{\mathscr{T}} &= [\mathbf{M}^\top \tilde{\mathbf{U}}_1]^\top (\mathbf{M}^{-1})(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}) \\ &= \tilde{\mathbf{U}}_1^\top \cancel{\mathbf{M}} \cancel{\mathbf{M}^{-1}} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}) \\ &= \tilde{\mathbf{U}}_1^\top (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}). \end{aligned} \tag{2.5.15}$$

As a result of (2.5.13) and (2.5.14), the gradient statistic is a quadratic form and $\xi_{\mathscr{T}} \overset{\mathscr{D}}{\to} \chi^2_q$. The advantage of (2.5.15) is that it does not depend on any kind of matrix, such as the Fisher information or observed information matrices.

# Chapter 3

# Generalised linear models with random effects

## 3.1 Introduction

The class of generalised linear models (GLMs) introduced by Nelder & Wedderburn (1972) established a new standard in statistical modelling. The GLMs extended the classic linear models for different situations where the response can be modelled by exponential family distributions and relating the response mean to the linear predictor through appropriate monotonic differentiable functions.

The concept of random effect modelling initially came up to accommodate subject-specific variability. More recently, this concept has been applied in situations where the model could not handle remain extra variability from the data. In this sense, the random effect is a part of the model assumed to be unknown, and can be regarded as a latent variable.

The GLMs with random effects considered is this thesis were proposed by Aitkin (1996b) for overdispersion modelling in GLMs and by Aitkin (1999) for variance components modelling. These models rely on the theory of finite mixture modelling which uses the EM algorithm for finding the maximum likelihood estimates proposed by Laird (1978). In the special case of a normally distributed random effect, Hinde (1982) proposed to employ tabulated Gauss-Hermite integration points and masses considering these values as constants.

## 3.2 The standard random effects model

Consider a generalised linear model with random effects (GLMwRE) for a data set containing $n$ independent observations of a response variable, given by $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, and corresponding observations on $p$ explanatory variables, given by $\boldsymbol{x}_i^\top = (x_{i1}, \ldots, x_{ip})^\top$, for $i = 1, \ldots, n$. The linear predictor for the $i$-th observation, $\eta_i$, has the form

$$\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + z_i^*, \tag{3.2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is the vector of regression parameters and $z_i^*$ is an unobserved random effect. The relationship between $y_i|z_i^*$ and $\eta_i$ is given by the conditional mean $\mu_i|z_i^* = \mathrm{E}[y_i|z_i^*]$ and the monotonic and differentiable *link function*, $g(\,\cdot\,)$ such that $\mu_i|z_i^* = g^{-1}(\eta_i)$.

By definition, $\boldsymbol{y}$ is a vector of independent random variables and each $y_i$, $i = 1, \ldots, n$ has a distribution in an exponential family with dispersion parameter. Thus, the probability density function of $y_i$ can be written as

$$f(y_i|\theta_i, \phi, z_i^*) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)], \tag{3.2.2}$$

where $\theta_1, \ldots, \theta_n$ are unknown parameters, $\phi > 0$ is a *precision* parameter common to all observations, and $b(\,\cdot\,)$ and $c(\,\cdot\,,\,\cdot\,)$ are known functions. The parameter estimation procedure requires the probability density function in (3.2.2) to be differentiable with respect to $\theta_i$ and $\phi$.

In (3.2.2), $\theta_i$ is related to $\mu_i|z_i^*$, and consequently to $\eta_i$, through two useful properties of an exponential family:

$$\mathrm{E}[y_i|z_i^*] = b'(\theta_i) \quad \text{and} \quad \mathrm{Var}[y_i|z_i^*] = \phi^{-1}V_i = \phi^{-1}b''(\theta_i), \tag{3.2.3}$$

where $V_i = V(\mu_i|z_i^*)$ and $V(\mu_i) = d\mu_i/d\theta_i = b''(\theta_i)$. The function $V(\mu_i|z_i^*)$ is called the *variance function* and $\phi^{-1}$ the *dispersion parameter*. Note that, unlike the GLM, $\mathrm{Var}[y_i] \geqslant \phi^{-1}V_i(\mu_i)$.

### 3.2.1 Random effects with normal distribution

According to Anderson & Hinde (1988) there are two approaches for the *unobserved* nature of the random effect $z_i^*$. The first consists in substituting $z_i^*$ by $\sigma z_i$ where $z_i \sim \mathcal{N}(0,1)$ and therefore, the linear predictor is writen as

$$\begin{aligned} \eta_i &= \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i \\ &= \dot{\boldsymbol{z}}_i^\top \boldsymbol{\gamma}. \end{aligned} \tag{3.2.4}$$

where $\dot{\boldsymbol{z}}_i = (\boldsymbol{x}_i^\top, z_i)^\top$ and $\boldsymbol{\gamma} = (\boldsymbol{\beta}^\top, \sigma)^\top$. The second is discussed in Subsection 3.2.2. The likelihood function for (3.2.4) is

$$\mathrm{L}^*(\boldsymbol{\gamma}, \phi) = \prod_{i=1}^{n} \int f(y_i|\boldsymbol{\gamma}, \phi, z_i)\varphi(z_i)dz_i \tag{3.2.5}$$

where $\varphi(\,\cdot\,)$ is the normal density and $f(\,\cdot\,)$ is the response density. However, the integral in (3.2.5) usually has no analytic solution. One of the several strategies suggested to solve this problem is approximation using a $K$-point *Gaussian quadrature* rule: for any function $h(z)$,

$$\int h(z)\varphi(z)\,dz \approx \sum_{k=1}^{K} \pi_k h(\tilde{z}_k)$$

where $\pi_k$ are the quadrature weights and $\tilde{z}_k$ the quadrature points. Both $\pi_k$ and $\tilde{z}_k$, $k = 1, \ldots, K$ are known and tabulated, see e.g. Golub & Welsch (1969) or Abramowitz & Stegun (1972).

Then the approximate likelihood is

$$\mathrm{L}^*(\boldsymbol{\gamma}, \phi) \approx \mathrm{L}(\boldsymbol{\gamma}, \phi) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k f(y_i|\boldsymbol{\gamma}, \phi, \tilde{z}_k) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k f_{ik}, \tag{3.2.6}$$

which is the likelihood for a per-observation $K$-component mixture of response distributions. According to Laird (1978), the approximation (3.2.6) becomes accurate already for a small integer $K$. Thus, in the subsequent theoretical development, we shall assume that this mixture model is in fact the true model so that $\mathrm{L}(\boldsymbol{\gamma}, \phi)$ is the true likelihood.

The choice of $K$ is arbitrary. For practical purposes, Einbeck & Hinde (2006b) suggests that the number of mass points $K$ should start with 1 and augmented as long as the likelihood improves.

### 3.2.2 Random effects with unspecified distribution

Restricting the distribution of the random effects to the normal distribution is the main disadvantage of the previous method. An alternative approach is to assume that $z_i^*$ in (3.2.4) has an unspecified density $\pi(\,\cdot\,)$. Hence, the likelihood for this model is

$$\mathrm{L}^*(\boldsymbol{\beta}, \phi) = \prod_{i=1}^{n} \int f(y_i|\boldsymbol{\beta}, \phi, z_i)\pi(z_i)dz_i. \tag{3.2.7}$$

Once again, for most choices of $\pi(\,\cdot\,)$ the integral of (3.2.7) cannot be calculated analytically. The solution proposed by Laird (1978) involves the approximation of the density $\pi(z_i)$ by a discrete distribution with an arbitrary number $K$ of *mass points* $z_k$ and $\tilde{\pi}_k$ *mass probabilities*, respectively, for $k = 1, \ldots, K$. Then, the integral in (3.2.7) is approximated by

$$\mathrm{L}^*(\boldsymbol{\beta}, \phi) \approx \mathrm{L}(\boldsymbol{\beta}, \phi, z_k) = \prod_{i=1}^{n} \sum_{k=1}^{K} f(y_i|\boldsymbol{\beta}, \phi, z_k)\tilde{\pi}_k = \prod_{i=1}^{n} \sum_{k=1}^{K} f_{ik}\tilde{\pi}_K. \tag{3.2.8}$$

where $f_{ik} = f(u_i|\boldsymbol{\beta}, \phi, z_k)$. The approximated likelihood in (3.2.8) corresponds to the model with linear predictor

$$\begin{aligned} \eta_{ik} &= \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{e}_{ik}^\top \boldsymbol{\zeta} \\ &= \dot{\boldsymbol{z}}_{ik}^\top \boldsymbol{\gamma}, \end{aligned} \tag{3.2.9}$$

with $z_i^* = \boldsymbol{e}_{ik}^\top \boldsymbol{\zeta}$ where $\boldsymbol{e}_{ik}$ is a $K$-dimensional vector of zeros except the one in the position $ik$, $\boldsymbol{\zeta} = (\zeta_1, \ldots \zeta_K)^\top$ is a vector of unknown parameters associated to the random effects, $\dot{\boldsymbol{z}}_{ik} = (\boldsymbol{x}_i^\top, \boldsymbol{e}_{ik}^\top)^\top$ and $\boldsymbol{\gamma} = (\boldsymbol{\beta}^\top, \boldsymbol{\zeta}^\top)^\top$ is the full vector of linear predictor parameters. Again, the choice of $K$ is arbitrary and the rule of thumb involves fit the model with $K = 1$ and then increase until the likelihood stabilises. An important practical advantage of this model in comparison to the Gaussian quadrature is that it does not restrict to one specific parametric model (the normal

distribution, for instance) for the random effects distribution. This means that the NPML model accommodates scenarios where the distribution of the random effects is asymmetric and discrete.

## 3.3 Unified notation and parameter estimation

Here we propose a general matrix notation for the GLMwRE. This notation is a formalisation of the implementation available in R package **npmlreg** (Einbeck et al., 2014). The notation is constructed so we can express the GLMwRE as a extension of the standard GLM and therefore extend some results of this model, such as the estimation procedure for the fixed effects.

Let $\ddot{\boldsymbol{y}}$ be a vector of $nK$ pseudo-observations

$$\ddot{\boldsymbol{y}} = (\underbrace{\boldsymbol{y}^\top, \boldsymbol{y}^\top, \ldots, \boldsymbol{y}^\top}_{K \text{ times}})^\top,$$

and $\ddot{\boldsymbol{z}}$ a vector of $nK$ mass points

$$\ddot{\boldsymbol{z}} = (\underbrace{\tilde{z}_1, \tilde{z}_1, \ldots, \tilde{z}_1}_{n \text{ times}}, \ldots, \underbrace{\tilde{z}_K, \tilde{z}_K, \ldots, \tilde{z}_K}_{n \text{ times}})^\top,$$

which will estimate the stacked vector of unobserved random effects.

The vector of expected values $\ddot{\boldsymbol{\mu}}$ is denoted by

$$\ddot{\boldsymbol{\mu}} = (\mu_{11}, \ldots, \mu_{n1}, \ldots, \mu_{1K}, \ldots, \mu_{nK})$$

where $\mu_{ik} = \mathrm{E}[y_i|\tilde{z}_k]$, for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Then, the linear predictor can be written as

$$g(\ddot{\boldsymbol{\mu}}) = \ddot{\boldsymbol{\eta}} = \ddot{\mathbf{Z}}\boldsymbol{\gamma} \tag{3.3.10}$$

where $g(\,\cdot\,)$ is the link function and

$$g(\ddot{\boldsymbol{\mu}}) = (g(\mu_{11}), \ldots, g(\mu_{n1}), \ldots, g(\mu_{1K}), \ldots, g(\mu_{nK}))^\top,$$
$$\ddot{\boldsymbol{\eta}} = (\eta_{11}, \ldots, \eta_{n1}, \ldots, \eta_{1K}, \ldots, \eta_{nK})^\top,$$

with $g(\mu_{ik}) = \eta_{ik}$, for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Finally, we define $\ddot{\dot{\mathbf{Z}}}$ as

$$\ddot{\dot{\mathbf{Z}}} = (\dot{\boldsymbol{z}}_{11}, \ldots, \dot{\boldsymbol{z}}_{n1}^{\top}, \ldots, \dot{\boldsymbol{z}}_{1K}, \ldots, \dot{\boldsymbol{z}}_{nK}). \tag{3.3.11}$$

We can consider (3.3.11) as a pseudo *model matrix* which includes the observed values of the covariates and the values to-be-estimated of the random effects. Then $\ddot{\dot{\mathbf{Z}}}$ is defined according to the chosen approach for the distribution of the random effects. For the model with normal random effects, $\ddot{\dot{\mathbf{Z}}}$ is a matrix with dimension $n \times p + 1$, where $\dot{\boldsymbol{z}}_{ik} = (\boldsymbol{x}_i^{\top}, \tilde{z}_k)^{\top}$ is for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. To match this model matrix, we have the vector of parameters $\boldsymbol{\gamma} = (\boldsymbol{\beta}^{\top}, \sigma)^{\top}$, with $\sigma > 0$. For random effects with unspecified distribution, $\ddot{\dot{\mathbf{Z}}}$ is a matrix with dimension $n \times p + K$, where $\dot{\boldsymbol{z}}_{ik} = (\boldsymbol{x}_i^{\top}, \boldsymbol{e}_{ik}^{\top})^{\top}$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Then, for this latter approach, the vector of parameters is $\boldsymbol{\gamma} = (\boldsymbol{\beta}^{\top}, \boldsymbol{\zeta}^{\top})^{\top}$. The log–likelihood function for the GLMwRE is

$$\ell(\gamma, \phi) = \log \mathrm{L}(\gamma, \phi) = \sum_{i=1}^{n} \log \left( \sum_{k=1}^{K} \pi_k f_{ik} \right), \tag{3.3.12}$$

and it turns out that equating the first partial derivatives to zero, that is $\partial \ell / \partial \boldsymbol{\gamma} = 0$, one obtains precisely the single–distribution score equations (Aitkin et al., 2009) for the GLM, but summed over $k = 1, \ldots, K$ and weighted by

$$\omega_{ik} = \frac{\pi_k f_{ik}}{\sum_{l=1}^{K} \pi_l f_{il}}. \tag{3.3.13}$$

Each $\omega_{ik}$ can be interpreted as the *posterior probability* that observation $y_i$ came from component $k$. Alternating between this estimation step and an update step for the $w_{ik}$ leads to an EM algorithm:

**E-step** Calculate weights $\omega_{ik}$ according to (3.3.13);

**M-step** Update the parameter estimates by fitting the GLM (3.3.10) with weights $\omega_{ik}$.

The ordinary generalised linear model (GLM) is a special case of the GLMwRE when $K = 1$ whether the choice of distribution for the random effects. In the normal random effects approach, the special case where $\sigma = 0$, the GLMwRE also reduces to an ordinary generalised linear model (GLM).

Inference for the precision parameter $\phi$, which we consider as a nuisance parameter, is not of primary interest in this paper. One can estimate $\phi$ in any EM iteration through

$$1/\hat{\phi} = \frac{1}{n}\sum_i \sum_k w_{ik} \frac{(y_i - \hat{\mu}_{ik})^2}{V(\hat{\mu}_{ik})},$$

using the current component mean estimates $\hat{\mu}_{ik} = g^{-1}(\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\sigma}\tilde{z}_k)$ and weights $w_{ik}$. The estimate $\hat{\phi}$ can be used at all occasions where $\phi$ appears henceforth in this manuscript. See for instance Einbeck & Hinde (2006a) for details.

For practical applications, it is very important to have reliable inferential tools for the regression parameters, $\boldsymbol{\beta}$. This is relevant, for instance, for the construction of confidence intervals or the assessment of strength of effects through hypothesis testing. Such inferences rely on the standard errors of the parameter estimates, $\hat{\boldsymbol{\beta}}$, which, in turn, can be computed via the Fisher information matrix. Therefore, the ability to compute this matrix accurately is paramount for most subsequent inferential procedures.

Writing the log–likelihood (3.3.12) as $\ell = \log \mathrm{L}(\boldsymbol{\gamma})$, the total score vector for $\boldsymbol{\gamma}$, $\mathbf{U} = \mathbf{U}(\boldsymbol{\gamma})$, is

$$\mathbf{U} = \frac{\partial \ell}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n} \frac{1}{\sum_{l=1}^{K} \pi_l f_{il}} \sum_{k=1}^{K} \pi_k \frac{\partial f_{ik}}{\partial \boldsymbol{\gamma}}.$$

By the chain rule, we find

$$\mathbf{U} = \sum_{i=1}^{n} \frac{\sum_{k=1}^{K} \pi_k f_{ik} \dfrac{\partial \log f_{ik}}{\partial \boldsymbol{\gamma}}}{\sum_{l=1}^{K} \pi_l f_{il}} = \phi \sum_{i=1}^{n} \sum_{k=1}^{K} \omega_{ik} \left\{ \frac{d\mu_{ik}}{d\eta_{ik}} \frac{(y_i - \mu_{ik})}{V_{ik}} \begin{pmatrix} \boldsymbol{x}_i \\ \tilde{z}_k \end{pmatrix} \right\}, \tag{3.3.14}$$

where $V_{ik} = V(\mu_{ik})$ and $\omega_{ik}$ is given by

$$\omega_{ik} = \frac{\pi_k f_{ik}}{\sum_{l=1}^{K} \pi_l f_{il}}. \tag{3.3.15}$$

Each $\omega_{ik}$ can be interpreted as the *posterior probability* that observation $y_i$ came from component $k$.

In matrix notation, we can rewrite $\mathbf{U}$ as

$$\mathbf{U} = \ddot{\mathbf{Z}}^{\top}\mathbf{D}(\ddot{\boldsymbol{y}} - \ddot{\boldsymbol{\mu}}), \tag{3.3.16}$$

where $\mathbf{D}$ is the diagonal matrix with diagonal entries $d_{11}, \ldots, d_{n1}, \ldots, d_{1K}, \ldots, d_{nK}$ given by

$$d_{ik} = \phi \frac{d\mu_{ik}}{d\eta_{ik}} \frac{\omega_{ik}}{V_{ik}}.$$

## 3.4 The variance components model

The variance components model is a generalisation of the standard random effects model for grouped data. We consider here the notation given by Aitkin (1999), where we have a two-stage random sample $y_{ij}$, where $i = 1, \ldots, n_j$ indexes the observations and $j = 1, \ldots, r$ the groups, with $\sum_{j=1}^{r} n_j = n$. Let denote $\mu_{ij}|z_j = \mathrm{E}[y_{ij}|z_j]$ the conditional mean of $y_{ij}$ given the unobserved random effect $z_j$. The mean is linked to a vector of $p$ covariates $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^{\top}$ by

$$g(\mu_{ij}|z_i) = \eta_{ij} = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + z_j, \quad \text{for} \quad i = 1, \ldots, n_j, \quad j = 1, \ldots r, \tag{3.4.17}$$

where $z_j$ is the $j$th random effect, $g(\,\cdot\,)$ is the link function and $\boldsymbol{\beta}$ is an unknown vector of parameters. We interpret $z_j$ as random intercepts for each group $j$. The likelihood is then defined by

$$\mathrm{L}^{*}(\boldsymbol{\beta}, \phi) = \prod_{j=1}^{r} \int \prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{\beta}, \phi, z_j)\pi(z_j)dz_j. \tag{3.4.18}$$

The integral in (3.4.18) has closed form just for the case where $z_j \sim \mathcal{N}(0, \sigma^2)$ which can be approximated by a Gaussian quadrature. Consequently we replace this integral by a finite sum over $K$ Gaussian quadrature mass points $z_k$ with means $\pi_k$ which is a similarly as the standard Gaussian quadrature model.

Similarly, one can consider that the $z_j$ has unspecified distribution and use the nonparametric estimation already stated for the NPML overdispersion model. The procedure is equivalent where the likelihood in (3.4.18) is than approximated by

a finite sum such as in (3.2.8) replacing $z_j$ and $\pi(z_j)$ by a discrete distribution with mass points $z_k$ and mass probabilities $\pi_k$ with $k = 1, \ldots, K$. The unknown parameters and $z_k$ and $\pi_k$ are estimated by EM algorithm.

### 3.4.1 The random coefficient model

Another variant is the *random coefficient model* which has a random slope $\beta_{1j} = \beta_1 + u_j$ where $u_j$ corresponds to a source of variation in regard to the mean of $\beta_{1j}$ so that $\mathrm{E}[u_j] = 0$. In this sense, the linear predictor can be expressed as

$$\begin{aligned} \eta_{ij} &= \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + u_j x_{1ij} + z_j \\ &= x_{ij}^{\top}\boldsymbol{\beta} + u_j x_{1ij} + z_j \end{aligned} \tag{3.4.19}$$

whilst marginally $u_j$ and $z_j$ have unknown joint distribution $\pi(z_j, u_j)$. The likelihood for (3.4.19) model is then defined as

$$\mathrm{L}^*(\boldsymbol{\beta}, \phi) = \prod_{j=1}^{r} \int \prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{\beta}, \phi, u_j, z_j)\pi(z_j, u_j)dz_j du_j \tag{3.4.20}$$

One can assume the distribution of $\pi(z, u)$ as a bivariate normal distribution with unknown covariance, which requires to solve numerically the integral in (3.4.20) over both $z_j$ and $u_j$. Still, Aitkin (1999, p. 120) note that this approach "[...]doubles the computational load[...]" and might be "[...]unusable for many random parameters[...]". An alternative solution is estimating the joint distribution of $z_j$ an $u_j$ nonparametrically, obtaining then the NPML estimate as a discrete distribution on finite number of mass-points ($\tilde{z}_k$,$\tilde{u}_k$) with mass probability $\tilde{\pi}_k$, for the $k$–th component.

## 3.5 Gradient test for GLMwRE

Consider testing

$$\begin{cases} \mathscr{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{(0)} \\ \mathscr{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^{(0)} \end{cases}$$

which induce the partitioning $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, where $\boldsymbol{\beta}_1$ is a $q$-dimensional vector of interest parameters and $\boldsymbol{\beta}_2$ is a $p-q$-dimensional vector of nuisance parameters with $q \leqslant p$. The corresponding partitioned model matrix is $\ddot{\ddot{\mathbf{Z}}} = (\ddot{\ddot{\mathbf{Z}}}_1, \ddot{\ddot{\mathbf{Z}}}_2)$.
The partitioning in $\boldsymbol{\beta}$ induces the following partition in the score vector

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\beta}|\boldsymbol{y}) \\ \mathbf{U}_2(\boldsymbol{\beta}|\boldsymbol{y}) \end{pmatrix} = \begin{pmatrix} \ddot{\ddot{\mathbf{Z}}}_1^\top \mathbf{D}(\ddot{\ddot{\boldsymbol{y}}} - \ddot{\ddot{\boldsymbol{\mu}}}) \\ \ddot{\ddot{\mathbf{Z}}}_2^\top \mathbf{D}(\ddot{\ddot{\boldsymbol{y}}} - \ddot{\ddot{\boldsymbol{\mu}}}) \end{pmatrix}$$

Thus, using the general definiton in (2.5.15) we express the gradient statistic for testing $\mathscr{H}_0$ for GLMwRE by

$$\xi_{\mathscr{T}} = \mathbf{U}_1(\tilde{\boldsymbol{\beta}}|\boldsymbol{y})^\top (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)}).$$

The likelihood ratio, Wald and Rao in the same context have the form, respectively,

$$\begin{aligned} \xi_{\mathscr{LR}} &= 2[\ell(\hat{\boldsymbol{\beta}}) - \ell(\tilde{\boldsymbol{\beta}})], \\ \xi_{\mathscr{W}} &= (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)})^\top \mathbf{K}_{11}(\hat{\boldsymbol{\beta}}|\boldsymbol{y})(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)}), \\ \xi_{\mathscr{R}} &= \mathbf{U}_1(\tilde{\boldsymbol{\beta}}|\boldsymbol{y})^\top \mathbf{K}^{11}(\tilde{\boldsymbol{\beta}}|\boldsymbol{y}) \mathbf{U}_1(\tilde{\boldsymbol{\beta}}|\boldsymbol{y}). \end{aligned}$$

where $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top)^\top$ and $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}_1^{(0)\top}, \tilde{\boldsymbol{\beta}}_2^\top)^\top$ are the maximum likelihood estimator under alternative and null hypothesis, respectively. Under $\mathscr{H}_0$ and for $n \to \infty$ the distribution of $\xi_{\mathscr{LR}}$, $\xi_{\mathscr{W}}$, $\xi_{\mathscr{R}}$ and $\xi_{\mathscr{T}}$ *should* have chi-square distribution with $q$ degrees of freedom.

# Chapter 4

# Fisher information matrix and standard errors

This chapter presents the formulae to compute the Fisher information matrix for the regression parameters of generalised linear models. The Fisher information matrix relies on the estimation of the response variance under the model assumptions. We propose two approaches to estimate the response variance: the first is based on an analytic formula (or a Taylor expansion for cases where we cannot obtain the closed-form) and the second is an approximation using the model estimates via the EM process. Further, simulations under several response distributions and a real data application involving a factorial experiment are presented and discussed. In terms of standard errors and coverage probabilities for model parameters, the proposed methods turn out to behave more reliably than the 'disparity rule' in (4.3.6) or approximations with the model fitted in the last EM iteration.

Despite the fact that the Fisher information matrix is not required for the gradient statistic, we made this effort to obtain the Wald and Rao statistic formulae. Thus allowing us to compare the gradient test properties to the likelihood ratio, Wald and Rao tests in Chapter 5.

## 4.1 The score vector and the Fisher information matrix

Recalling the notation present in Section 3.3, we have the log–likelihood (3.3.12) as $\ell = \log L(\boldsymbol{\gamma})$, the total score vector for $\boldsymbol{\gamma}$, $\mathbf{U} = \mathbf{U}(\boldsymbol{\gamma})$, is

$$\mathbf{U} = \frac{\partial \ell}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n} \frac{1}{\sum_{l=1}^{K} \pi_l f_{il}} \sum_{k=1}^{K} \pi_k \frac{\partial f_{ik}}{\partial \boldsymbol{\gamma}}.$$

By the logarithmic differentiation, we find

$$\mathbf{U} = \sum_{i=1}^{n} \frac{\sum_{k=1}^{K} \pi_k f_{ik} \frac{\partial \log f_{ik}}{\partial \boldsymbol{\gamma}}}{\sum_{l=1}^{K} \pi_l f_{il}} = \phi \sum_{i=1}^{n} \sum_{k=1}^{K} \omega_{ik} \left\{ \frac{d\mu_{ik}}{d\eta_{ik}} \frac{(y_i - \mu_{ik})}{V_{ik}} \begin{pmatrix} \boldsymbol{x}_i \\ \tilde{z}_k \end{pmatrix} \right\}, \tag{4.1.1}$$

where $\omega_{ik}$ is given by (3.3.15) and $V_{ik} = V(\mu_{ik})$. In matrix notation, we can rewrite $\mathbf{U}$ as

$$\mathbf{U} = \ddot{\mathbf{Z}}^{\top} \mathbf{D} (\ddot{\boldsymbol{y}} - \ddot{\boldsymbol{\mu}}), \tag{4.1.2}$$

where $\mathbf{D}$ is the diagonal matrix with diagonal entries $d_{11}, \ldots, d_{n1}, \ldots, d_{1K}, \ldots, d_{nK}$ given by

$$d_{ik} = \phi \frac{d\mu_{ik}}{d\eta_{ik}} \frac{\omega_{ik}}{V_{ik}}.$$

Similarly, denote by $\mathbf{K} = \mathbf{K}(\boldsymbol{\gamma})$ the GLMwRE Fisher information matrix for $\boldsymbol{\gamma}$. Then $\mathbf{K} = \mathrm{Var}\,[\mathbf{U}]$ and, from (4.1.2), we have

$$\begin{aligned} \mathbf{K} &= \mathrm{Var}\left[\ddot{\mathbf{Z}}^{\top} \mathbf{D} (\ddot{\boldsymbol{y}} - \ddot{\boldsymbol{\mu}})\right] \\ &= \ddot{\mathbf{Z}}^{\top} \mathbf{D} \mathrm{Var}\,[\ddot{\boldsymbol{y}}]\, \mathbf{D} \ddot{\mathbf{Z}} \\ &= \ddot{\mathbf{Z}}^{\top} \mathbf{D} \ddot{\boldsymbol{\Upsilon}} \mathbf{D} \ddot{\mathbf{Z}}, \end{aligned}$$

where $\ddot{\boldsymbol{\Upsilon}} = \mathrm{Var}\,[\ddot{\boldsymbol{y}}]$ is the unconditional variance-covariance matrix for $\ddot{\boldsymbol{y}}$. Since the

observations in the GLMwRE are independent,

$$\begin{aligned}\mathrm{Cov}(y_i, y_i) &= \mathrm{Var}(y_i), \quad \forall i \in \{1, \dots n\}, \text{ and} \\ \mathrm{Cov}(y_i, y_j) &= 0, \quad \forall i \neq j \in \{1, \dots n\},\end{aligned}$$

one finds for the $K$ copies in $\ddot{\boldsymbol{y}}$ that

$$\begin{aligned}\mathrm{Cov}(y_i^{(k)}, y_i^{(l)}) &= \mathrm{Var}(y_i), \quad \forall i \in \{1, \dots n\}, k, l \in \{1, \dots K\}, \text{ and} \\ \mathrm{Cov}(y_i^{(k)}, y_j^{(l)}) &= 0 \quad \forall i \neq j \in \{1, \dots n\}, k, l \in \{1, \dots K\},\end{aligned}$$

where $y_i^{(k)}$ and $y_i^{(l)}$ are the $k$th and $l$th copies of $y_i$, respectively for $\forall i \in \{1, \dots n\}$ and $k, l \in \{1, \dots K\}$.

Therefore,

$$\ddot{\boldsymbol{\Upsilon}} = \underbrace{\begin{pmatrix} \boldsymbol{\Upsilon} & \boldsymbol{\Upsilon} & \cdots & \boldsymbol{\Upsilon} \\ \boldsymbol{\Upsilon} & \boldsymbol{\Upsilon} & \cdots & \boldsymbol{\Upsilon} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Upsilon} & \boldsymbol{\Upsilon} & \cdots & \boldsymbol{\Upsilon} \end{pmatrix}}_{K \text{ times}} \Bigg\} K \text{ times},$$

where $\boldsymbol{\Upsilon} = \mathrm{diag}(\upsilon_i)$ and $\upsilon_i = \mathrm{Var}(y_i)$. For compactness of the notation, let $\boldsymbol{\Psi} = \mathbf{D}\ddot{\boldsymbol{\Upsilon}}\mathbf{D}$. The Fisher information matrix for the GLMwRE is then

$$\mathbf{K} = \ddot{\mathbf{Z}}^\top \boldsymbol{\Psi} \ddot{\mathbf{Z}}.$$

Here, the response variances play an important role and the following Section 4.2 develops the necessary formulae.

## 4.2 Response variance

Recall that, in model (3.2.4), the $z_i$ follow a standard normal distribution. That is, though they are approximated by a discrete set $\tilde{z}_1, \dots \tilde{z}_K$ for estimation purposes,

they are random in nature, so that the unconditional mean and variance of $y_i$ are

$$\mathrm{E}[y_i] = \mathrm{E}[\mathrm{E}[y_i|z_i]] = \mathrm{E}[\mu_i] \tag{4.2.3}$$

and

$$\begin{aligned}\mathrm{Var}(y_i) &= \mathrm{E}[\mathrm{Var}[y_i|z_i]] + \mathrm{Var}[\mathrm{E}[y_i|z_i]] \\ &= \phi^{-1}\mathrm{E}[V(\mu_i)] + \mathrm{Var}[\mu_i].\end{aligned} \tag{4.2.4}$$

The remaining task is to determine $\mathrm{E}[V(\mu_i)]$ and $\mathrm{Var}[\mu_i]$. This can be achieved either approximately, by use of the Gaussian quadrature rule, or analytically, based on explicit expressions depending on the response distribution and link function. We explain both approaches below.

### 4.2.1 Estimation via analytic expressions

We derived the analytic form of $\mathrm{E}[V(\mu_i)]$ and $\mathrm{Var}[\mu_i]$ for Normal, Gamma, Poisson, Binomial and Inverse Gaussian response distribution and a wide range of commonly used link functions. The resulting expressions for $\mathrm{Var}(y_i)$ are summarized in Table 4.1. Some combinations of distribution and link function required the use of a Taylor expansion, which is indicated by a $\circledast$. All such expansions were made to third order. Of course, for the practical use in (4.2.3) and (4.2.4), $\boldsymbol{\beta}$ and $\sigma$ need to be replaced by their corresponding estimates.

All derivations are give in detail in Appendix B but here we explain two exemplary situations. Firstly, suppose a GLMwRE with normal response with Gaussian random effects. Consider that the identity link function is appropriate for this case. Thus, we have

$$\mu_i = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i$$

and $V(\mu) = 1$. Therefore, $\mathrm{E}[y_i] = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ and the response variance is

$$\begin{aligned}\mathrm{Var}(y_i) &= \phi^{-1}\mathrm{E}[1] + \mathrm{Var}[\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i] \\ &= \phi^{-1} + \sigma^2.\end{aligned}$$

However, there are cases in which there is no analytical solution for $\mathrm{E}[V(\mu_i)]$ and $\mathrm{Var}[\mu_i]$. In such cases, an approximate solution can be obtained by expanding $V(\mu_i)$ and $\mu_i$ by Taylor series around $z_i = 0$. Therefore, secondly, consider a GLMwRE with Gaussian random effects, Gamma response and inverse link. For this configuration, $V(\mu) = \mu^2$ and

$$\mu_i = \frac{1}{\eta_i} = \frac{1}{\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i}.$$

Thus,

$$\begin{aligned} \mathrm{E}[\phi^{-1} V(\mu_i)] &= \phi^{-1} \mathrm{E}\left[(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i)^{-2}\right], \quad \text{and} \\ \mathrm{Var}[\mu_i] &= \mathrm{Var}[(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i)^{-1}]. \end{aligned}$$

By Taylor expansion around 0, we have

$$(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i)^{-1} \approx (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-1} - (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-2} \sigma z_i + (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-3} \sigma^2 z_i^2 - (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4} \sigma^3 z_i^3,$$

and

$$(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i)^{-2} \approx (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-2} - 2(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-3} \sigma z_i + 3(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4} \sigma^2 z_i^2 - 4(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-5} \sigma^3 z_i^3.$$

Therefore, after some algebra, we have the response variance as

$$\begin{aligned} \mathrm{Var}(y_i) \approx \phi^{-1} &\left[(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-2} + 3(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4} \sigma^2\right] + \\ &+ (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4} \sigma^2 + 8(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-6} \sigma^4 + 15(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-8} \sigma^6. \end{aligned}$$

For practical purposes, the Taylor expansions presented in Table 4.1 are sufficiently accurate for any $\sigma > 0$ and $\boldsymbol{x}_i^\top \boldsymbol{\beta}$ such that $|\boldsymbol{x}_i^\top \boldsymbol{\beta}| > \sigma$ and $|\sigma / (\boldsymbol{x}_i^\top \boldsymbol{\beta})| < 0.4$.

### 4.2.2 Estimation via Gaussian Quadrature

Approximating

$$\mathrm{E}[V(\mu_i)] \approx \sum_{k=1}^{K} V_{ik} \pi_k, \qquad \mathrm{Var}[\mu_i] = \sum_{k=1}^{K} \mu_{ik}^2 \pi_k - \left(\sum_{k=1}^{K} \mu_{ik} \pi_k\right)^2,$$

Table 4.1: Variance of response under Gaussian quadrature models.

| Response Distribution | Link function | $\mathrm{Var}(y_i)$ |
|---|---|---|
| Normal | identity | $\phi^{-1}+\sigma^2$ |
| | log | $\phi^{-1}+\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})+\sigma^2\}(\exp\{\sigma^2\}-1)$ |
| | inverse$^{\circledast}$ | $\phi^{-1}+(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-4}\sigma^2+8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-6}\sigma^4+15(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-8}\sigma^6$ |
| Gamma | identity | $(\phi^{-1}+1)\sigma^2+\phi^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2$ |
| | log | $\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})+\sigma^2\}[(\phi^{-1}+1)\exp\{\sigma^2\}-1]$ |
| | inverse$^{\circledast}$ | $\phi^{-1}\left[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-2}+3(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-4}\sigma^2\right]+$ $+(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-4}\sigma^2+8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-6}\sigma^4+15(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-8}\sigma^6$ |
| Poisson | log | $(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\exp\{\sigma^2/2\}\times$ $\times[1+(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\exp\{\sigma^2/2\}(\exp\{\sigma^2\}-1)]$ |
| | identity | $\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma^2$ |
| | square root | $(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2+4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^2+\sigma^2+2\sigma^4$ |
| Binomial | logit$^{\circledast}$ | $\frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1}-\frac{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^2}-$ $-\frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^3}+$ $+\frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^3-(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2]\sigma^2}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^4}-$ $-\frac{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]^2\sigma^4}{4(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^6}$ |
| | probit$^{\circledast}$ | $\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{2}-\Phi^2(\boldsymbol{x}_i^\top\boldsymbol{\beta})+$ $+(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})-$ $-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^4\phi^2(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{4}$ |
| | cauchit$^{\circledast}$ | $\frac{1}{4}-\frac{1}{\pi^2}\left\{\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta})-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2}{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2+1]^2}\right\}^2$ |
| | log | $\exp\left\{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\frac{\sigma^2}{2}\right\}-\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})+\sigma^2\}$ |
| | comp. log-log$^{\circledast}$ | $\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}+\frac{\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}}-$ $-\exp\{-2\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}-\frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{\exp\{2\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}}-$ $-\frac{[\exp\{4(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}-2\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}+\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}]\sigma^4}{4\exp\{2\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}}$ |
| Inv. Gaussian | $1/\mu^{2\circledast}$ | $\phi^{-1}\left[\frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}+\frac{15\sigma^2}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}\right]+$ $+\frac{\sigma^2}{4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3}+\frac{\sigma^4}{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5}+\frac{375\sigma^6}{256(\boldsymbol{x}_i^\top\boldsymbol{\beta})^7}$ |
| | inverse$^{\circledast}$ | $\phi^{-1}\left[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-3}+6(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-5}\sigma^2\right]+$ $+(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-4}\sigma^2+8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-6}\sigma^4+15(\boldsymbol{x}_i^\top\boldsymbol{\beta})^{-8}\sigma^6$ |
| | identity | $\phi^{-1}[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3+3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2]+\sigma^2$ |
| | log | $\phi^{-1}\exp\left\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})+\frac{9\sigma^2}{2}\right\}+\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma^2)\}-$ $-\exp\left\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})+\frac{\sigma^2}{2}\right\}$ |

$^{\circledast}$ Approximated via Taylor expansion.

one has

$$\mathrm{Var}(y_i) \approx \phi^{-1} \sum_{k=1}^{K} V_{ik}\pi_k + \sum_{k=1}^{K} \mu_{ik}^2 \pi_k - \left( \sum_{k=1}^{K} \mu_{ik}\pi_k \right)^2 . \tag{4.2.5}$$

The number of mass points $K$ as the quantities $\phi$, $V_{ik}$, $\mu_{ik}$ and $\pi_k$ are obtained from the fitted model.

An advantage of this expression is that it extends to nonparametric maximum likelihood estimation (NPML) of random effect models (Aitkin et al., 2009) by substituting $\mu_k$ and $z_k$ with their estimates from the final EM iteration. In the context of Gaussian quadrature, which is the focus of this manuscript, we found (4.2.5) to behave very similarly to the analytic expressions above, as demonstrated in the following section.

## 4.3 Examples

We now provide two examples, using simulated and real data, to illustrate the use of the Fisher information matrix for the computation of standard errors of the regression parameter estimates. The first example involves four simulated data sets based on 90 observations, one each for models with Poisson, Gamma, Normal and Inverse Gaussian responses. The second example illustrates the application of the Inverse Gaussian distribution to a real dataset with 30 observations. The results in each example can be reproduced with code available in the supplementary material. As reference, we use the standard errors obtained by two procedures: (i) via Monte Carlo with 10.000 replicates, (ii) via the heuristic formula

$$\mathrm{se}(\hat{\beta}_j) \doteq \frac{|\hat{\beta}_j|}{\sqrt{\Delta\,\mathrm{disp}_j}}, \tag{4.3.6}$$

where $\Delta\,\mathrm{disp}_j$ is the change in disparity ($-2\ell$) when omitting the explanatory variable $x_j$ (Aitkin et al., 2009, pg. 439). A natural limitation of this formula is that it is not possible to compute the standard error for $\sigma$ or the intercept term. Therefore the results for $\sigma$ and $\beta_0$ are left blank. The values given in column (iii) are the standard errors of $\hat{\boldsymbol{\gamma}}$ in the GLM fit of the last EM iteration.

The results (iv) and (v) are the standard errors obtained using the analytic formula

for variance (or the Taylor expansion) from Section 4.2.1 and the approximation from Section 4.2.2, respectively. In all of (i) to (v), the actual model fitting was carried out using R function `alldist` Einbeck et al. (2014), using $K = 3$ throughout. For comparative purposes, we also provide the standard errors (vi) and parameter estimates $\hat{\boldsymbol{\gamma}}^*$ produced by the `glmer` function (Bates et al., 2014), using the default option for argument `nAGQ` which implies a Laplace Approximation for the integral in (3.2.5). The `glmer` does not return the standard errors for $\sigma$ thus the results are left blank.

### 4.3.1 Simulated data example

For each case, we simulate 10000 data sets of size $n = 90$ based on the following linear predictor

$$\eta_i = \beta_0 + \beta_1 x_i + \beta_{2i} + \sigma z_i, \quad i = 1, \ldots, n,$$

with the intercept $\beta_0 = 1$ for Poisson, Gamma and Normal cases and $\beta_0 = 1.5$ for Inverse Gaussian. The covariate $x$ is generated from $\mathcal{U}(0, 1)$ with coefficient $\beta_1 = -1$ for Poisson, Gamma and Normal cases and $\beta_1 = -0.125$ for Inverse Gaussian. The $\beta_{2i}$ represent the coefficients of a factor with three levels, which $\beta_{2i} = (i \mod 3) - 1$ for Poisson, Gamma and Normal cases and $\beta_{2i} = 0.125 \times \{(i \mod 3) - 1\}$ for Inverse Gaussian. The random effect term is generated from $\mathcal{N}(0, 1)$ and the amount of variability due to the random effects is controlled by $\sigma$ with value 0.125 for all models. We choose $\tau$ equal 1 for Gaussian and Gamma model, and equal 1/64 for the Inverse Gaussian model. The link functions are log for Poisson and Gamma, identity for Normal and inverse for Inverse Gaussian. We opt for a different set of parameter values for the Inverse Gaussian model due to the inverse link constraint $\eta_i \neq 0$ and the larger value of $\tau$ offers a balance for $\mu_i^3$ in $\mathrm{Var}(y_i) = \mathrm{E}[\phi^{-1}\mu_i^3] + \mathrm{Var}[\mu_i]$. We resample a new dataset for cases where `alldist` or `glmer` did not fit the model. For the Normal model, `lmer` is used instead of `glmer`.

Tables 4.2, 4.3, 4.4 and 4.5 display, respectively, the average value of $\hat{\boldsymbol{\gamma}}$ as well as the average standard errors of $\hat{\boldsymbol{\gamma}}$ for models fitted to the simulated response distributions Poisson, Gamma, Normal and Inverse Gaussian.

Table 4.2: Estimated fixed effects and respective standard errors (Poisson model with log link)

| | | | se($\hat{\boldsymbol{\gamma}}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\boldsymbol{\gamma}$ | $\overline{\hat{\boldsymbol{\gamma}}}$ | (i) | (ii) | (iii) | (iv) | (v) | (vi) | $\overline{\hat{\boldsymbol{\gamma}}}^*$ |
| $\beta_0$ | 1 | 0.98608 | 0.17855 | — | 0.17658 | 0.17149 | 0.17658 | 0.18070 | 0.98629 |
| $\beta_1$ | -1 | -1.00175 | 0.24900 | 0.23640 | 0.24903 | 0.24019 | 0.24903 | 0.25527 | -1.00199 |
| $\beta_{22}$ | 1 | 1.00395 | 0.16835 | 0.16769 | 0.16607 | 0.16157 | 0.16607 | 0.16924 | 1.00445 |
| $\beta_{23}$ | -1 | -1.01731 | 0.27644 | 0.23448 | 0.27395 | 0.27098 | 0.27396 | 0.27595 | -1.01726 |
| $\sigma$ | 0.125 | 0.12582 | 0.07255 | — | 0.07192 | 0.06912 | 0.07192 | — | 0.08803 |

Standard errors for $\hat{\boldsymbol{\gamma}}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with analytic variance;
(v) Fisher information matrix using approximation (4.2.5); and
(vi) Laplace approximation (`glmer` output).
* shows the estimates for $\boldsymbol{\gamma}$ obtained via `glmer`.

For the standard errors of the regression parameters, we see from columns (iv) and (v) of all four tables that the values obtained using our proposed methods are slightly below those obtained by Monte Carlo resampling (i). The standard errors (ii) using the disparity rule offer numbers close to (i) in the Poisson, Gamma and Normal examples. However, (ii) shows rather small standard error estimates for the Inverse Gaussian example. The standard errors (iii) taken from the generalised linear model fit of the last EM iteration are quite accurate for the Poisson model, but are underestimating the true standard error for the Gamma, Normal and Inverse Gaussian models. We did not observe much difference between the approaches (iv) and (v) using the Fisher information, though the standard errors using (v) were slightly more accurate in general, especially for the Inverse Gauss scenario where a Taylor expansion was used for the analytic formula (iv). The standard errors using `glmer` were usually higher than than those of (i), (iv) and (v), except for the inverse Gaussian model, and were reasonably consistent with overall results. However, it is observed that `glmer` struggles to estimate the $\sigma$ parameter correctly, sometimes underestimating (Poisson model) but mostly overestimating, and does not provide a value for the standard error of $\hat{\sigma}$ at all. We further note that, for the Gamma model, the average of `glmer` estimates for $\beta_0$ is less than half of the true value, which might indicate an identifiability issue.

For the study of coverage probabilities, we provide results estimated for the Poisson and Gamma models in Tables 4.6 and 4.7, respectively. On each table, the numbers show the results of estimated coverage probability (C.P.) computed through confi-

Table 4.3: Estimated fixed effects and respective standard errors (Gamma model with log link)

| | | | se($\hat{\boldsymbol{\gamma}}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\boldsymbol{\gamma}$ | $\overline{\hat{\boldsymbol{\gamma}}}$ | (i) | (ii) | (iii) | (iv) | (v) | (vi) | $\overline{\hat{\boldsymbol{\gamma}}}^*$ |
| $\beta_0$ | 1 | 0.96764 | 0.27011 | — | 0.14468 | 0.23456 | 0.24826 | 0.29173 | 0.40759 |
| $\beta_1$ | -1 | -0.99366 | 0.38621 | 0.33437 | 0.20773 | 0.33435 | 0.35646 | 0.41886 | -0.98002 |
| $\beta_{22}$ | 1 | 0.99877 | 0.26766 | 0.26031 | 0.14359 | 0.23107 | 0.24639 | 0.28946 | 1.00816 |
| $\beta_{23}$ | -1 | -0.99716 | 0.26561 | 0.25669 | 0.14350 | 0.23092 | 0.24624 | 0.28927 | -0.98762 |
| $\sigma$ | 0.125 | 0.12427 | 0.11265 | — | 0.05980 | 0.09665 | 0.10261 | — | 1.25017 |

Standard errors for $\hat{\boldsymbol{\gamma}}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with analytic variance;
(v) Fisher information matrix using approximation (4.2.5); and
(vi) Laplace approximation (`glmer` output).
* shows the estimates for $\boldsymbol{\gamma}$ obtained via `glmer`.

Table 4.4: Estimated fixed effects and respective standard errors (Normal model with identity link)

| | | | se($\hat{\boldsymbol{\gamma}}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\boldsymbol{\gamma}$ | $\overline{\hat{\boldsymbol{\gamma}}}$ | (i) | (ii) | (iii) | (iv) | (v) | (vi) | $\overline{\hat{\boldsymbol{\gamma}}}^*$ |
| $\beta_0$ | 1 | 0.99425 | 0.27235 | — | 0.15096 | 0.25531 | 0.25903 | 0.26715 | 0.99440 |
| $\beta_1$ | -1 | -0.99892 | 0.38068 | 0.32482 | 0.21190 | 0.35837 | 0.36360 | 0.37497 | -0.99938 |
| $\beta_{22}$ | 1 | 1.00644 | 0.26113 | 0.24809 | 0.14772 | 0.24983 | 0.25348 | 0.26142 | 1.00670 |
| $\beta_{23}$ | -1 | -0.99233 | 0.26064 | 0.24702 | 0.14786 | 0.25007 | 0.25372 | 0.26171 | -0.99250 |
| $\sigma$ | 0.125 | 0.12676 | 0.11043 | — | 0.06150 | 0.10400 | 0.10553 | — | 0.79139 |

Standard errors for $\hat{\boldsymbol{\gamma}}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with analytic variance;
(v) Fisher information matrix using approximation (4.2.5); and
(vi) Laplace approximation (`lmer` output).
* shows the estimates for $\boldsymbol{\gamma}$ obtained via `lmer`.

dence intervals which use the standard error estimates (i), (ii), (iii), (iv), (v) and (vi) already discussed. Our intention here is to show two rather different scenarios, where the first (Poisson model) exemplifies well behaved numbers of C.P. and, in the second (Gamma model), an extreme case where we are able to note an evident contrast between the methods on the C.P.s.

Assuming that an specific method to compute the standard errors is reasonably good to compute the confidence intervals, we overall expect values close to the usual true confidence levels (C.L.) of 90%, 95% and 99% on average. Thus, we observe that for the Poisson model on Table 4.6, all five methods are acceptable according to our criteria, except for the disparity rule in (ii). However, for the Gamma model on Table 4.7, we note that the Monte-Carlo values in (i) are rather close to the true confidence levels, followed by the estimates via Fisher information matrix in (v) and (iv). The values computed using the disparity rule in (ii), the last EM iteration in

Table 4.5: Estimated fixed effects and respective standard errors (Inv. Gaussian model with inverse link)

| | | | se($\hat{\boldsymbol{\gamma}}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\boldsymbol{\gamma}$ | $\overline{\hat{\boldsymbol{\gamma}}}$ | (i) | (ii) | (iii) | (iv) | (v) | (vi) | $\overline{\hat{\boldsymbol{\gamma}}}^*$ |
| $\beta_0$ | 1.5 | 1.50068 | 0.03923 | — | 0.02197 | 0.02801 | 0.03771 | 0.02686 | 1.54060 |
| $\beta_1$ | -0.125 | -0.12495 | 0.05492 | 0.01756 | 0.03127 | 0.03961 | 0.05366 | 0.03140 | -0.12421 |
| $\beta_{22}$ | 0.125 | 0.12508 | 0.03986 | 0.01699 | 0.02243 | 0.02879 | 0.03849 | 0.03743 | 0.12406 |
| $\beta_{23}$ | -0.125 | -0.12523 | 0.03848 | 0.01702 | 0.02153 | 0.02702 | 0.03694 | 0.04122 | -0.12429 |
| $\sigma$ | 0.125 | 0.12490 | 0.01615 | — | 0.00910 | 0.01166 | 0.01562 | — | 0.19321 |

Standard errors for $\hat{\boldsymbol{\gamma}}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with Taylor expansion of the analytic variance;
(v) Fisher information matrix using approximation (4.2.5); and
(vi) Laplace approximation (`glmer` output).
* shows the estimates for $\boldsymbol{\gamma}$ obtained via `glmer`.

Table 4.6: Estimated coverage probabilities (Poisson model with log link)

| | | C.P. (%) | | | | |
|---|---|---|---|---|---|---|
| | C.L. (%) | $\beta_0 = 1$ | $\beta_1 = -1$ | $\beta_{22} = 1$ | $\beta_{23} = -1$ | $\sigma = 0.125$ |
| (i) | 90.00 | 89.96 | 90.26 | 90.19 | 90.30 | 89.94 |
| | 95.00 | 94.86 | 94.94 | 94.97 | 94.89 | 94.93 |
| | 99.00 | 98.90 | 98.97 | 98.86 | 98.69 | 98.94 |
| (ii) | 90.00 | — | 87.30 | 89.72 | 83.85 | — |
| | 95.00 | — | 92.67 | 94.56 | 90.26 | — |
| | 99.00 | — | 97.11 | 98.48 | 96.06 | — |
| (iii) | 90.00 | 89.86 | 90.46 | 89.96 | 90.87 | 89.80 |
| | 95.00 | 94.91 | 95.19 | 94.88 | 95.71 | 95.05 |
| | 99.00 | 99.05 | 99.11 | 98.90 | 99.17 | 99.08 |
| (iv) | 90.00 | 88.87 | 89.12 | 88.96 | 90.48 | 87.71 |
| | 95.00 | 94.07 | 94.31 | 94.23 | 95.44 | 93.21 |
| | 99.00 | 98.80 | 98.77 | 98.66 | 99.09 | 98.26 |
| (v) | 90.00 | 89.86 | 90.46 | 89.96 | 90.87 | 89.80 |
| | 95.00 | 94.91 | 95.19 | 94.88 | 95.71 | 95.05 |
| | 99.00 | 99.05 | 99.11 | 98.90 | 99.17 | 99.08 |
| (vi) | 90.00 | 89.80 | 90.46 | 90.03 | 90.86 | — |
| | 95.00 | 94.93 | 94.92 | 94.85 | 95.69 | — |
| | 99.00 | 99.00 | 99.08 | 98.83 | 99.18 | — |

Estimated coverage probabilities of the CI for $\hat{\boldsymbol{\gamma}}$ computed using the standard errors obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with analytic variance;
(v) Fisher information matrix using approximation (4.2.5); and
(vi) `glmer` output.

(iii) and, especially, from `glmer` output in (vi) are overall smaller than the confidence levels.

Table 4.7: Estimated coverage probabilities (Gamma model with log link)

| | | C.P. (%) | | | | |
|---|---|---|---|---|---|---|
| | C.L. (%) | $\beta_0 = 1$ | $\beta_1 = -1$ | $\beta_{22} = 1$ | $\beta_{23} = -1$ | $\sigma = 0.125$ |
| (i) | 90.00 | 90.20 | 90.10 | 89.88 | 89.88 | 90.05 |
| | 95.00 | 94.88 | 94.74 | 94.96 | 94.89 | 94.73 |
| | 99.00 | 98.60 | 98.99 | 99.07 | 98.95 | 98.73 |
| (ii) | 90.00 | — | 80.33 | 86.35 | 86.58 | — |
| | 95.00 | — | 85.53 | 91.62 | 91.45 | — |
| | 99.00 | — | 91.05 | 96.36 | 96.08 | — |
| (iii) | 90.00 | 62.39 | 62.56 | 62.43 | 62.46 | 61.78 |
| | 95.00 | 70.70 | 70.94 | 70.59 | 71.07 | 70.77 |
| | 99.00 | 83.03 | 83.34 | 82.49 | 83.12 | 83.02 |
| (iv) | 90.00 | 84.06 | 84.13 | 83.66 | 84.13 | 83.40 |
| | 95.00 | 90.41 | 90.48 | 90.22 | 90.29 | 89.32 |
| | 99.00 | 96.52 | 96.80 | 96.89 | 96.72 | 95.15 |
| (v) | 90.00 | 86.63 | 87.04 | 86.38 | 86.94 | 86.74 |
| | 95.00 | 92.42 | 92.54 | 92.33 | 92.54 | 92.31 |
| | 99.00 | 97.65 | 97.92 | 97.93 | 98.03 | 97.76 |
| (vi) | 90.00 | 39.44 | 79.00 | 78.44 | 79.17 | — |
| | 95.00 | 50.37 | 83.99 | 83.35 | 83.90 | — |
| | 99.00 | 68.89 | 87.75 | 87.80 | 88.04 | — |

Coverage probabilities of the CI for $\hat{\boldsymbol{\gamma}}$ computed using the standard errors obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with Taylor expansion of the analytic variance;;
(v) Fisher information matrix using approximation (4.2.5); and
(vi) `glmer` output.

### 4.3.2 Real data example

As a real data example, we take a subsample of the data from a $5 \times 2$ factorial experiment given by Ostle & Mensing (1963). This subsample is provided in the `R` library `mdscore` (da Silva-Júnior et al., 2014), using the syntax `data(strength)`. It is of interest to investigate how the impact strength of an insulating material is affected by the lot (I, II, III, IV, V) of the material and the type of specimen cut (lengthwise and crosswise). Previous analysis of the original dataset is given in Shuster & Miura (1972) and for a subsample in da Silva-Júnior et al. (2014). In our analysis, we assume that the impact strength measurements of a given replicate corresponding to the $i$-th cut and $j$–th lot are independently distributed as inverse Gaussian distributions with means $\mu_{ij}$ and a fixed dispersion parameter. We jus-

tify this choice mainly because the variable is strictly positive. Suppose the linear predictor in the inverse link scale corresponds to the two–way interaction model

$$\mu_{ij}^{-1} = \tau_0 + \tau_i + \beta_j + (\tau\beta)_{ij} + \sigma z, \quad i = 1, 2, \ j = 1, 2, ..., 5, \tag{4.3.7}$$

where $\tau_1 = 0$, $\beta_1 = 0$, $(\tau\beta)_{11} = \cdots = (\tau\beta)_{15} = (\tau\beta)_{21} = 0$, and $z$, is a random effect that has Gaussian distribution.

Again, the estimate $\hat{\boldsymbol{\gamma}}$ was obtained using `alldist`, and columns (ii) to (v) of Table 4.8 report the standard errors of $\hat{\boldsymbol{\gamma}}$ obtained using the different techniques. Additionally, column (i) reports Monte-Carlo standard errors for $\hat{\boldsymbol{\gamma}}$ by generating 10000 new samples of size 30 responses based on (4.3.7), taking $\hat{\boldsymbol{\gamma}}$ as "true" parameter values, and refitting the model for each one. It is further noted that, for this data set and model specification, the `glmer` attempt to fitting the model (4.3.7) failed to converge in our trials even when we relax the tolerances and the algorithm stopping criteria.

Table 4.8: Estimated fixed effects and respective standard errors (strength data da Silva-Júnior et al. (2014))

| | | se($\hat{\boldsymbol{\gamma}}$) | | | | |
|---|---|---|---|---|---|---|
| | $\hat{\boldsymbol{\gamma}}$ | (i) | (ii) | (iii) | (iv) | (v) |
| $\tau_0$ | 1.01704 | 0.07042 | — | 0.03197 | 0.06869 | 0.06832 |
| $\tau_2$ | 0.32828 | 0.10564 | 0.11340 | 0.04873 | 0.10462 | 0.10413 |
| $\beta_2$ | 0.03201 | 0.10043 | 0.09876 | 0.04557 | 0.09780 | 0.09728 |
| $\beta_3$ | 0.35915 | 0.10711 | 0.11543 | 0.04904 | 0.10531 | 0.10482 |
| $\beta_4$ | 0.14128 | 0.10273 | 0.10293 | 0.04676 | 0.10037 | 0.09986 |
| $\beta_5$ | 0.82348 | 0.11757 | 0.15159 | 0.05359 | 0.11513 | 0.11468 |
| $(\tau\beta)_{22}$ | -0.40636 | 0.14657 | 0.15279 | 0.06637 | 0.14247 | 0.14175 |
| $(\tau\beta)_{23}$ | -0.10864 | 0.15825 | 0.15968 | 0.07322 | 0.15726 | 0.15661 |
| $(\tau\beta)_{24}$ | -0.35020 | 0.14937 | 0.15481 | 0.06841 | 0.14689 | 0.14619 |
| $(\tau\beta)_{25}$ | -0.19501 | 0.17043 | 0.17270 | 0.07879 | 0.16928 | 0.16867 |
| $\sigma$ | 0.00887 | 0.02119 | — | 0.01131 | 0.37348 | 0.37174 |

Standard errors for $\hat{\boldsymbol{\gamma}}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with Taylor expansion of the analytic variance; and
(v) Fisher information matrix using approximation (4.2.5).

The numbers in (iv) and (v), for the fixed effects, are very slightly smaller than their

counterparts in (i) and (ii). However, and contrary to our simulations presented on Subsection 4.3.1, the numbers in (iv) and (v) for $\hat{\sigma}$ are rather large. This might be due to misspecification of the random effects distribution. Finally, the numbers in (iii) are considerably smaller than their counterparts in (i), (ii), (iv) and (v).

# Chapter 5

# Simulated data experiments and real data examples

We present here a set of simulation experiments to evaluate the performance of the gradient test for nested GLMwREs and four illustrative applications to real data sets. The properties investigated are the type I error, estimated by the test rejection rates — the proportion of simulated replicas for which the null hypothesis is rejected considering that this hypothesis is true — and the power of the test, which is estimated by the test rejection rates under a Pitman sequence of local alternative hypotheses. The computation has been performed using R code provided in Appendix A.

The simulation experiment shown in Section 5.1 aims to study the properties of the gradient test in generalised linear models with random effects, here called GQ, NPML and VC models for *Gaussian quadrature*, *Nonparametric maximum likelihood* and *variance components*, respectively. The first two models are commonly applied for small and moderate overdispersion which means that the fixed effects plays a major role for explaining the model response variability. The latter model is a generalisation of the NPML model and it is often used for grouped data, also admitting random slopes in its formulae. For stability purposes, we discard the large overdispersion scenario as it might lead to identifiability issues. Here we consider large overdispersion when half or more of the true variability of the model is explained by the random effect. We assume the normal distribution with fixed variance for GQ

models and a discrete distribution for the random effects simulated for both NPML and VC models. We also perform the likelihood ratio, the Wald and the Rao tests for the same simulated samples as comparative measure.
Finally we present four real data examples with applications for the gradient test in Section 5.2.

## 5.1 Simulated data experiments

### 5.1.1 General design

We consider the following GLMwREs for the simulation study

$$\eta_i^{\text{GQ}} = \beta_0 + \beta_1 x_{1i} + \beta_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \sigma z_i, \text{ for } i = 1, \ldots, n \tag{5.1.1}$$

$$\eta_i^{\text{NPML}} = \beta_1 x_{1i} + \beta_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + z_i, \text{ for } i = 1, \ldots, n \tag{5.1.2}$$

$$\eta_{ij}^{\text{VC}} = (\beta_1 + u_{ij}) x_{1ij} + \beta_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + z_{ij}, \text{ for } i = 1, \ldots, n_j, \ j = 1, \ldots, 10 \tag{5.1.3}$$

where $\eta_i^{\text{GQ}}$, $\eta_i^{\text{NPML}}$ and $\eta_i^{\text{VC}}$ are the linear predictors for the GQ, NPML and VC fittings, respectively. In (5.1.1) and (5.1.2), $x_{1i}$, $x_{3i}$ and $x_{4i}$ are samples of size $n$ from $\mathcal{U}(0,1)$, $\mathcal{F}(2,5)$ and $t(3)$, respectively. The same applies to (5.1.3), however each $x_{\cdot ij}$ has size $n_j$ such that $\sum_{j=1}^{10} n_j = n$. The random effect $z_i$ in (5.1.1) is a sample from a standard normal distribution and $\sigma = 8^{-1}$. Analogously, $z_i$ in (5.1.2) is sampled from a discrete distribution which takes $K$ values from $\mathcal{N}(1, 8^{-2})$ (or $\mathcal{N}(2, 8^{-2})$ for inverse Gaussian) and probabilities from $\mathcal{U}(0,1)$. In (5.1.3), $u_{ij}$ is sampled from a discrete distribution with $K$ values taken from $\mathcal{N}(0, 8^{-2})$ and probabilities from $\mathcal{U}(0,1)$; $z_{ij}$ is sampled from a discrete distribution with $K$ values from $\mathcal{N}(1, 8^{-1})$ (or $\mathcal{N}(2, 8^{-2})$ for inverse Gaussian) and probabilities from $\mathcal{U}(0,1)$. For the GQ models, the parameter values are $\beta_0 = 1$, $\beta_1 = -1$ and $\beta_{2i} = (i \mod 3) - 1$ except for the inverse Gaussian response model where $\beta_0 = 2$, $\beta_1 = 1$ and $\beta_{2i} = (i \mod 3)$. For the NPML models, $\beta_1 = -1$ and $\beta_{2i} = (i \mod 3) - 1$ except for the inverse Gaussian response model where $\beta_1 = 1$ and $\beta_{2i} = (i \mod 3)$. For VC models $\beta_1 = -1$ and $\beta_{2i} = (i \mod 3) - 1$ except for the inverse Gaussian

response model where $\beta_1 = 1$ and $\beta_{2i} = (i \mod 3)$. This means that $\eta_i^{\text{GQ}}$, $\eta_i^{\text{NPML}}$ and $\eta_{ij}^{\text{VC}}$ are equivalent on average within the same response distribution.

Consider the scenario that $\boldsymbol{\beta}^{\text{GQ}} = (\beta_0, \ldots, \beta_4)^\top$, $\boldsymbol{\beta}^{\text{NPML}} = (\beta_1, \ldots, \beta_4)^\top$ and $\boldsymbol{\beta}^{\text{VC}} = (\beta_1, \ldots, \beta_4)^\top$ the full vectors of fixed effects parameters for GQ, NPML and VC models, respectively. Therefore, (5.1.1), (5.1.2) and (5.1.3) can be also expressed as

$$\eta_i^{\text{GQ}} = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{\text{GQ}} + \sigma z_i, \text{ for } i = 1, \ldots, n \tag{5.1.4}$$

$$\eta_i^{\text{NPML}} = \boldsymbol{x}_{\bar{i}}^\top \boldsymbol{\beta}^{\text{NPML}} + z_i, \text{ for } i = 1, \ldots, n \tag{5.1.5}$$

$$\eta_{ij}^{\text{VC}} = \boldsymbol{x}_{\bar{ij}}^\top \boldsymbol{\beta}^{\text{VC}} + z_i, \text{ for } i = 1, \ldots, n_j, \ j = 1, \ldots, 10. \tag{5.1.6}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_{\bar{i}}$ are the vectors of covariates with and without intercept, respectively. Our aim here is to test the composite hypotheses

$$\begin{cases} \mathscr{H}_0 : \boldsymbol{\beta}_1 = \mathbf{0} \\ \mathscr{H}_1 : \boldsymbol{\beta}_1 \neq \mathbf{0} \end{cases} \tag{5.1.7}$$

where $\boldsymbol{\beta}_1 = (\beta_3, \beta_4)^\top$, hence $\boldsymbol{\beta}^\bullet = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^{\bullet\top})^\top$ and $\boldsymbol{\beta}_2^{\text{GQ}} = (\beta_0, \beta_1, \beta_2)^\top$ and $\boldsymbol{\beta}_2^{\text{NPML}} = (\beta_1, \beta_2)^\top$ for GQ and NPML models, respectively. This also leads to $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}^\top, \boldsymbol{x}_{i2}^{\bullet\top})^\top$ where $\boldsymbol{x}_{i1} = (x_{i3}, x_{i4})^\top$ and $\boldsymbol{x}_{i2}^{\bullet\top}$ can either be $\boldsymbol{x}_{i2} = (1, x_{i1}, x_{i2})^\top$ or $\boldsymbol{x}_{\bar{i}2} = (x_{i1}, x_{i2})^\top$. On top of that, $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)^\top$ and $\tilde{\boldsymbol{\beta}} = (\mathbf{0}, \tilde{\boldsymbol{\beta}}_2^{\bullet\top})^\top$ are the unrestricted and restricted to $\mathscr{H}_0$ maximum likelihood estimators for $\boldsymbol{\beta}$, respectively. From now on, every notation which shows the accents $\wedge$ or $\sim$ will refer to the correspondingly unrestricted and restricted estimators.

For broad comprehension of the problem, we take samples of $n = 50$, 100, 200 and 400 observations, which covers more or less the "small" to the "large" spectrum of sample sizes. We set $K = 3, 5$ and 7 for GQ and NPML and $K = 3$ and 5 for VC models. The number of replicas for each set-up of the model is 10000. The procedure automatically discarded any replica where any of the test statistics was not strictly positive and the experiment continued until 10000 valid replications were obtained.

### 5.1.2 Size and power properties

The size of a test can be defined as the probability of rejecting the null hypothesis when this hypothesis is true. In other words, it is the probability of making a Type I error. Therefore, for this simulation experiment, the data is generated under the null hypothesis and we investigate the empirical rejection rates of each test compared to the common nominal levels of 10%, 5% and 1%. The rejection rates are reported. The power of a test is the probability of rejecting the null hypothesis when this hypothesis is false. For the purpose of simulation, the data is generated under local alternatives

$$\begin{cases} \mathscr{H}_0 : \boldsymbol{\beta}_1 = 0 \\ \mathscr{H}_1 : \boldsymbol{\beta}_1 = \dfrac{\delta}{\sqrt{n}} \tilde{\text{se}}(\tilde{\boldsymbol{\beta}}_1) \end{cases} \tag{5.1.8}$$

where $\delta$ take values in a numeric sequence of 51 equidistant numbers in the interval $[-4, 4]$ and $\tilde{\text{se}}(\tilde{\boldsymbol{\beta}}_1)$ is estimated in a secondary Monte Carlo simulation with 10000 replicates for each scenario. Tables 5.1, 5.2, 5.3, 5.4 and 5.5 show the Monte Carlo estimated standard errors for $\tilde{\boldsymbol{\beta}}_1 = (\tilde{\beta}_3, \tilde{\beta}_4)^\top$ and for binomial, Poisson, gamma, normal and inverse Gaussian, respectively. The alternative hypothesis $\mathscr{H}_1$ in (5.1.8) is called Pitman sequence of local alternative hypotheses and converges to the null hypothesis at rate $n^{-1/2}$ (Peers, 1971; Hayakawa, 1975).

Table 5.1: Monte Carlo standard errors for $\tilde{\beta}_3$ and $\tilde{\beta}_4$ for binomial models

| | | GQ model | | NPML model | | VC model | |
|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ |
| 50 | 3 | 1.46415 | 1.82727 | 2.19798 | 2.54094 | 2.91587 | 3.36982 |
| 50 | 5 | 1.44961 | 1.87723 | 2.85978 | 3.16544 | 2.83188 | 3.02434 |
| 50 | 7 | 1.47694 | 1.79628 | 2.98358 | 3.26368 | | |
| 100 | 3 | 1.13778 | 1.55799 | 1.50800 | 2.04084 | 2.56241 | 2.92467 |
| 100 | 5 | 1.04691 | 1.59621 | 1.52408 | 2.13323 | 2.05865 | 2.71025 |
| 100 | 7 | 1.11610 | 1.57240 | 1.43595 | 2.11622 | | |
| 200 | 3 | 0.96699 | 1.44651 | 1.18572 | 1.74985 | 1.99913 | 2.43245 |
| 200 | 5 | 0.94696 | 1.44095 | 1.13571 | 1.78531 | 2.08107 | 2.51722 |
| 200 | 7 | 0.92400 | 1.42501 | 1.19805 | 1.78480 | | |
| 400 | 3 | 0.85493 | 1.34797 | 1.00412 | 1.65738 | 2.10897 | 2.57456 |
| 400 | 5 | 0.80302 | 1.36962 | 1.03307 | 1.65864 | 1.87447 | 2.62991 |
| 400 | 7 | 0.82325 | 1.34317 | 0.97941 | 1.60115 | | |

Table 5.2: Monte Carlo standard errors for $\tilde{\beta}_3$ and $\tilde{\beta}_4$ for Poisson models

| | | GQ model | | NPML model | | VC model | |
|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ |
| 50 | 3 | 0.36106 | 0.44640 | 0.33815 | 0.45852 | 0.33815 | 0.45852 |
| 50 | 5 | 0.33471 | 0.44559 | 0.32971 | 0.48655 | 0.32971 | 0.48655 |
| 50 | 7 | 0.37912 | 0.43213 | 0.32620 | 0.44730 | | |
| 100 | 3 | 0.29309 | 0.38531 | 0.28260 | 0.43399 | 0.28260 | 0.43399 |
| 100 | 5 | 0.29597 | 0.37754 | 0.27553 | 0.42399 | 0.27553 | 0.42399 |
| 100 | 7 | 0.27925 | 0.37614 | 0.27318 | 0.42965 | | |
| 200 | 3 | 0.24227 | 0.35176 | 0.24850 | 0.41673 | 0.24850 | 0.41673 |
| 200 | 5 | 0.25221 | 0.34898 | 0.25666 | 0.40721 | 0.25666 | 0.40721 |
| 200 | 7 | 0.23924 | 0.34078 | 0.24540 | 0.41933 | | |
| 400 | 3 | 0.20216 | 0.32865 | 0.23649 | 0.39591 | 0.23649 | 0.39591 |
| 400 | 5 | 0.21631 | 0.33249 | 0.24052 | 0.38981 | 0.24052 | 0.38981 |
| 400 | 7 | 0.20962 | 0.32838 | 0.24538 | 0.38708 | | |

### 5.1.3 Results

Tables 5.7, 5.8, 5.9, 5.10 and 5.11 show the null rejection rates of each of the four tests for Poisson, binomial, gamma, normal and inverse Gaussian responses. Table 5.6 shows the Figure enumerations for the tests power plots for each scenario of our simulation.

First, the overall perception is that the gradient test null rejection rates are closer to the nominal levels, except for a few cases. As expected for an asymptotic test applied to a finite sample, the numbers for the smallest sample size $n = 50$ are far from the true nominal levels. However these improve gradually as $n$ gets to 400. Nevertheless, the other three tests perform even worse for $n = 50$. Still, the rejection rates of all tests are better for GQ and VC models than for NPML models, perhaps because in the latter, the number of nuisance parameters increases as $K$ increases. This phenomena can be observed as $K$ increases, the changes in the rejection rates in the GQ and VC models are mostly due to Monte Carlo variability in contrast to what happens in the NPML model.

The rejection rates for the Poisson response can be found in Table 5.7. We see that the numbers for the Rao and Wald test are far from the nominal levels but with a different behaviour according to the type of the model. For the GQ model, the

Table 5.3: Monte Carlo standard errors for $\tilde{\beta}_3$ and $\tilde{\beta}_4$ for gamma models

| | | GQ model | | NPML model | | VC model | |
|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ |
| 50 | 3 | 0.48525 | 0.72672 | 0.57586 | 0.90532 | 0.93090 | 1.23926 |
| 50 | 5 | 0.49069 | 0.71173 | 0.61317 | 0.87460 | 0.88350 | 1.10735 |
| 50 | 7 | 0.51879 | 0.71193 | 0.56898 | 0.89708 | | |
| 100 | 3 | 0.42852 | 0.63736 | 0.53638 | 0.77311 | 0.82672 | 1.04117 |
| 100 | 5 | 0.40386 | 0.64392 | 0.50369 | 0.87129 | 0.87403 | 1.12312 |
| 100 | 7 | 0.37654 | 0.66677 | 0.54549 | 0.85204 | | |
| 200 | 3 | 0.36303 | 0.60578 | 0.41191 | 0.67903 | 1.00759 | 1.07008 |
| 200 | 5 | 0.38373 | 0.64668 | 0.43820 | 0.74200 | 0.80962 | 1.09490 |
| 200 | 7 | 0.37316 | 0.60675 | 0.44970 | 0.78838 | | |
| 400 | 3 | 0.35983 | 0.60604 | 0.37529 | 0.62296 | 0.91902 | 1.03648 |
| 400 | 5 | 0.36530 | 0.60388 | 0.37057 | 0.66236 | 0.90707 | 1.19654 |
| 400 | 7 | 0.35117 | 0.59884 | 0.39887 | 0.69159 | | |

Rao test is *conservative*, rejecting less than the nominal levels and the Wald test has opposite behaviour, i.e., being *liberal*. Nevertheless, for the NPML model, the two tests change roles where the Rao test shows *liberal* numbers and the Wald test presents *conservative* numbers. The gradient test and the likelihood ratio test show similar rejection rates. But, the gradient test is less sensible to the variation of $K$ on NPML and VC models where it has rejection rates closer to the nominal levels.

The binomial response model rejection rates are presented in Table 5.8. The GQ model results show that the Wald test is *conservative* where the other are *liberal*. In this context, the Rao test numbers are less distant to the nominal levels. For the NPML model, the numbers for all four tests are far from the nominal levels and the likelihood ratio test showed slightly better performance. However this minor advantage does not imply that the likelihood ratio test is reliable enough and we cannot recommend any of the four tests for this scenario. For VC models, the numbers show that the likelihood ratio test is preferable for smallest sample sizes and the gradient test for larger ones. The gradient test also showed numbers less sensitive to the increase in $K$ value.

On the numbers for the gamma response model shown in Table 5.9, we notice that the Rao test for GQ models are closer to the nominal levels, followed by those from the likelihood ratio, gradient and Wald tests, in this order. In contrast, the numbers

Table 5.4: Monte Carlo standard errors for $\tilde{\beta}_3$ and $\tilde{\beta}_4$ for normal models

| | | GQ model | | NPML model | | VC model | |
|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ |
| 50 | 3 | 0.40644 | 0.68685 | 0.51928 | 0.85372 | 1.00440 | 1.15088 |
| 50 | 5 | 0.42502 | 0.69523 | 0.48911 | 0.84988 | 0.89231 | 1.18465 |
| 50 | 7 | 0.41263 | 0.66861 | 0.50860 | 0.81681 | | |
| 100 | 3 | 0.38163 | 0.61902 | 0.48246 | 0.72444 | 1.06387 | 1.19559 |
| 100 | 5 | 0.40598 | 0.63617 | 0.46686 | 0.81638 | 1.04281 | 1.05629 |
| 100 | 7 | 0.39859 | 0.61461 | 0.46244 | 0.81717 | | |
| 200 | 3 | 0.37632 | 0.59516 | 0.38870 | 0.65897 | 0.85826 | 1.12505 |
| 200 | 5 | 0.37480 | 0.60374 | 0.42959 | 0.77349 | 0.92799 | 1.18951 |
| 200 | 7 | 0.38679 | 0.58329 | 0.42460 | 0.81239 | | |
| 400 | 3 | 0.35200 | 0.58634 | 0.34679 | 0.60326 | 0.82824 | 1.02110 |
| 400 | 5 | 0.34661 | 0.59011 | 0.37570 | 0.66561 | 1.06520 | 1.23060 |
| 400 | 7 | 0.33279 | 0.58839 | 0.40109 | 0.75397 | | |

for NPML model show that the Rao test is quite *conservative* and the Wald test is quite *liberal*. In the same case, the likelihood ratio test is less liberal than the Wald test but still far from the nominal levels. Here, the gradient test is also *liberal* however clearly closer to the nominal levels. On VC models, the two simulated tests have numbers much closer to the nominal levels with a slight advantage to the likelihood ratio test.

The rejection rates for normal response models are presented in Table 5.10. First, the Rao and gradient test numbers are exactly the same for the GQ model. We believe that this happens because, for this particular case, the Rao and gradient statistics are essentially the same in formulae. The numbers of the likelihood ratio and Wald test are more *conservative* than the Rao/gradient test in this case. For the NPML model, we noticed that the Wald and Rao test are still highly sensitive to the increase value of $K$. The likelihood ratio and the gradient test has numbers closer to but the latter is less sensitive to the increase in $K$ and shows faster convergence to the nominal levels. The same behaviour is observed on the VC model for this case.

Lastly, the results in Table 5.11 refers to the models with inverse Gaussian response. For the GQ model, all tests are *liberal* but the gradient test shows numbers approaching to the nominal levels. In this sense, the Rao test is the second best, followed

Table 5.5: Monte Carlo standard errors for $\tilde{\beta}_3$ and $\tilde{\beta}_4$ for inverse Gaussian models

| | | GQ model | | NPML model | | VC model | |
|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ | $\tilde{\text{se}}(\tilde{\beta}_3)$ | $\tilde{\text{se}}(\tilde{\beta}_4)$ |
| 50 | 3 | 0.24190 | 0.39387 | 0.33015 | 0.50010 | 0.72335 | 0.64688 |
| 50 | 5 | 0.27986 | 0.39884 | 0.32808 | 0.51103 | 0.83312 | 0.62830 |
| 50 | 7 | 0.26162 | 0.39837 | 0.31281 | 0.49875 | | |
| 100 | 3 | 0.24343 | 0.37792 | 0.27228 | 0.43595 | 0.77675 | 0.63333 |
| 100 | 5 | 0.24587 | 0.36857 | 0.28961 | 0.47781 | 0.77969 | 0.60643 |
| 100 | 7 | 0.25272 | 0.37537 | 0.30900 | 0.50034 | | |
| 200 | 3 | 0.21888 | 0.36603 | 0.21700 | 0.37698 | 0.72816 | 0.66840 |
| 200 | 5 | 0.21944 | 0.36389 | 0.26255 | 0.44618 | 0.83758 | 0.65518 |
| 200 | 7 | 0.22421 | 0.34814 | 0.27679 | 0.46221 | | |
| 400 | 3 | 0.21515 | 0.35625 | 0.21203 | 0.36429 | 0.71384 | 0.65915 |
| 400 | 5 | 0.20396 | 0.35608 | 0.22003 | 0.36857 | 0.84287 | 0.63724 |
| 400 | 7 | 0.20580 | 0.35013 | 0.22017 | 0.41587 | | |

Table 5.6: Table of Figure enumerations for non-null rejection curves for each simulated scenario

| Response | $K$ | GQ model | NPML model | VC model |
|---|---|---|---|---|
| Poisson | 3 | 5.1 | 5.4 | 5.7 |
| | 5 | 5.2 | 5.5 | 5.8 |
| | 7 | 5.3 | 5.6 | |
| binomial | 3 | 5.9 | 5.12 | 5.15 |
| | 5 | 5.10 | 5.13 | 5.16 |
| | 7 | 5.11 | 5.14 | |
| gamma | 3 | 5.17 | 5.20 | 5.23 |
| | 5 | 5.18 | 5.21 | 5.24 |
| | 7 | 5.19 | 5.22 | |
| normal | 3 | 5.25 | 5.28 | 5.31 |
| | 5 | 5.26 | 5.29 | 5.32 |
| | 7 | 5.27 | 5.30 | |
| inverse Gaussian | 3 | 5.33 | 5.36 | 5.39 |
| | 5 | 5.34 | 5.37 | 5.40 |
| | 7 | 5.35 | 5.38 | |

by the likelihood ratio and Wald tests. At a first look, one could think that the Rao test has rejection rates closer to the nominal levels. However, those numbers are not consistent for all $K$ and $n$ values. In fact, the Rao test numbers deviate from the nominal levels when the sample size $n$ increases. Alternatively, the other three tests does not suffer from this behaviour and the gradient test numbers show

faster convergence to the nominal levels. Moreover, the gradient test numbers are less sensible to the variation in $K$.

We have some remarks regarding to the non-null rejection rates represented by the power plots referenced in Figures 5.1 to 5.40. First, for GQ models we notice that there is no visible difference of changing $K$ within the response distribution. This can be seen in Figures 5.1, 5.2 and 5.3 which present the power curves for Poisson GQ models with $K = 3, 5$ and 7 respectively.

Second, the power curves show some convergence as long as the sample size $n$ increases, for instance in Figures 5.17, 5.18 and 5.19 which show the power curves for gamma GQ models with $K = 3, 5$ and 7, respectively. Our criteria of convergence relies on how close is the bottom of the curve (the region around $\delta = 0$) to the true nominal level. In this sense, we observe that the convergence of the power for the Wald and Rao tests is slightly slower, for instance in Figure 5.11.

On the other hand, the likelihood ratio and gradient test show quite similar power curves, sometimes one cannot distinguish the difference between the two, for instance in Figures 5.25, 5.26 and 5.27 which show the power curves for normal GQ models with $K = 3, 5$ and 7, respectively. This behaviour is also shown in VC model results as we can see in Figures 5.31 and 5.32. On the other hand, the Wald and Rao tests show curves rather different as we can see, for instance, in the Figures 5.2, 5.17 and 5.27. When we read the simulation results of type I error and power together for GQ models we can conclude that the gradient test has rejection numbers close to the nominal levels without loosing much power in comparison to the other tests.

One interesting aspect of the results for NPML model is that the small difference seen in the GQ model between the four tests is amplified, e.g. in Figures 5.20, 5.21 and 5.22. We also see a much slower convergence on the Wald and Rao test curves when the sample size $n$ increases, see for instance Figures 5.28, 5.29 and 5.30. On the other hand, the likelihood ratio and gradient test show some improvement when the sample size $n$ increases, for instance in Figures 5.36, 5.37 and 5.38. We also notice that the Wald and Rao tests are quite sensitive to the increase in $K$ showing a much slower convergence for larger $K$, for example in Figures 5.12, 5.13 and 5.14. In contrast, the likelihood ratio and the gradient test are less sensitive to

the variation in $K$. The gradient test is sometimes slightly less powerful then the likelihood ratio test, such as in Figures 5.12 and 5.13. However, for most of the cases the gradient test is most powerful then the likelihood ratio test, e.g. in Figures 5.36, 5.29 and 5.19.

Finally, we notice some small difference between the likelihood ratio and the gradient test on VC models for the smallest sample size $n = 50$, for example in Figures 5.23, 5.24, 5.31 and 5.32. Apart from the smallest sample size scenario, the difference of the two test curves is almost negligible. The likelihood ratio and the gradient test power curves are not much affected by the variation in $K$, as we can see in Figures 5.39 and 5.40. Therefore the gradient test is at least equivalent to the likelihood ratio test in power for VC models.

Table 5.7: Null rejection rates of the four tests for Poisson models

| | | GQ model | | | | NPML model | | | | VC model | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{T}}$ |
| 50 | 3 | 10.11 | 11.92 | 7.90 | 10.48 | 9.36 | 4.89 | 16.50 | 8.91 | 11.69 | 12.15 |
| 50 | 3 | 5.01 | 6.70 | 3.86 | 5.46 | 4.50 | 2.24 | 9.64 | 4.04 | 5.78 | 6.21 |
| 50 | 3 | 1.13 | 1.74 | 0.83 | 1.40 | 0.73 | 0.42 | 2.82 | 0.60 | 1.44 | 1.63 |
| 100 | 3 | 10.32 | 12.15 | 8.56 | 10.50 | 9.98 | 5.31 | 16.78 | 9.57 | 10.53 | 10.73 |
| 100 | 3 | 5.20 | 6.51 | 4.18 | 5.43 | 4.97 | 2.49 | 10.06 | 4.75 | 5.39 | 5.53 |
| 100 | 3 | 1.15 | 1.65 | 0.78 | 1.34 | 0.88 | 0.46 | 3.28 | 0.92 | 1.20 | 1.29 |
| 200 | 3 | 10.45 | 11.77 | 8.53 | 10.72 | 10.06 | 5.59 | 17.77 | 9.88 | 10.71 | 10.85 |
| 200 | 3 | 4.98 | 6.20 | 4.17 | 5.22 | 5.05 | 2.65 | 10.80 | 4.88 | 5.24 | 5.37 |
| 200 | 3 | 0.95 | 1.47 | 0.74 | 1.12 | 0.93 | 0.52 | 3.25 | 1.01 | 1.32 | 1.34 |
| 400 | 3 | 9.68 | 11.15 | 8.25 | 9.82 | 9.50 | 5.06 | 16.52 | 9.68 | 10.29 | 10.27 |
| 400 | 3 | 4.86 | 5.93 | 4.23 | 4.97 | 4.64 | 2.13 | 9.89 | 4.64 | 5.28 | 5.26 |
| 400 | 3 | 0.97 | 1.41 | 0.77 | 1.04 | 0.87 | 0.46 | 2.83 | 0.93 | 1.06 | 1.03 |
| 50 | 5 | 10.59 | 12.61 | 8.28 | 10.98 | 9.20 | 2.68 | 25.46 | 8.47 | 11.25 | 11.65 |
| 50 | 5 | 5.07 | 7.05 | 3.91 | 5.52 | 4.66 | 1.13 | 17.08 | 4.12 | 5.73 | 6.18 |
| 50 | 5 | 1.00 | 1.98 | 0.86 | 1.20 | 0.77 | 0.16 | 6.79 | 0.71 | 1.40 | 1.57 |
| 100 | 5 | 9.98 | 11.63 | 8.24 | 10.25 | 8.98 | 2.51 | 25.16 | 8.61 | 10.82 | 11.01 |
| 100 | 5 | 4.98 | 6.47 | 3.99 | 5.38 | 4.37 | 1.12 | 16.74 | 4.19 | 5.53 | 5.58 |
| 100 | 5 | 0.95 | 1.49 | 0.78 | 1.09 | 0.83 | 0.20 | 6.36 | 0.89 | 1.14 | 1.25 |
| 200 | 5 | 9.57 | 11.00 | 8.19 | 9.94 | 9.05 | 2.21 | 25.90 | 8.96 | 11.02 | 11.06 |
| 200 | 5 | 4.89 | 5.86 | 4.02 | 5.20 | 4.32 | 0.99 | 17.24 | 4.23 | 5.55 | 5.64 |
| 200 | 5 | 1.11 | 1.55 | 0.88 | 1.21 | 0.77 | 0.23 | 6.86 | 0.70 | 1.10 | 1.14 |
| 400 | 5 | 9.92 | 11.42 | 8.36 | 10.09 | 9.61 | 2.39 | 26.39 | 9.56 | 10.88 | 10.76 |
| 400 | 5 | 4.97 | 6.15 | 4.00 | 5.08 | 4.74 | 1.05 | 17.68 | 4.61 | 5.79 | 5.75 |
| 400 | 5 | 0.85 | 1.51 | 0.72 | 1.01 | 0.91 | 0.18 | 7.18 | 0.92 | 1.13 | 1.12 |
| 50 | 7 | 10.19 | 12.32 | 8.04 | 10.61 | 9.60 | 1.28 | 34.85 | 8.95 | | |
| 50 | 7 | 5.14 | 6.81 | 4.05 | 5.52 | 4.74 | 0.43 | 25.11 | 4.25 | | |
| 50 | 7 | 1.27 | 1.89 | 0.79 | 1.55 | 0.69 | 0.02 | 11.86 | 0.55 | | |
| 100 | 7 | 9.96 | 11.79 | 8.32 | 10.16 | 8.80 | 1.39 | 33.75 | 8.42 | | |
| 100 | 7 | 5.00 | 6.44 | 3.93 | 5.12 | 4.33 | 0.60 | 24.08 | 4.04 | | |
| 100 | 7 | 1.01 | 1.80 | 0.84 | 1.09 | 0.79 | 0.08 | 11.05 | 0.73 | | |
| 200 | 7 | 9.67 | 11.25 | 8.52 | 9.93 | 9.21 | 1.03 | 34.57 | 9.10 | | |
| 200 | 7 | 4.93 | 6.05 | 3.96 | 5.21 | 4.38 | 0.41 | 25.02 | 4.10 | | |
| 200 | 7 | 0.96 | 1.46 | 0.84 | 1.00 | 0.69 | 0.03 | 12.33 | 0.75 | | |
| 400 | 7 | 10.15 | 11.77 | 8.66 | 10.23 | 9.37 | 1.21 | 34.08 | 9.32 | | |
| 400 | 7 | 5.19 | 6.24 | 4.44 | 5.35 | 4.71 | 0.54 | 25.17 | 4.69 | | |
| 400 | 7 | 0.93 | 1.54 | 0.76 | 1.07 | 1.07 | 0.16 | 12.24 | 1.00 | | |

Table 5.8: Null rejection rates of the four tests for binomial models

| | | GQ model | | | | NPML model | | | | VC model | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{T}}$ |
| 50 | 3 | 16.52 | 8.91 | 11.03 | 21.44 | 40.86 | 43.83 | 29.02 | 39.03 | 19.36 | 23.38 |
| 50 | 3 | 9.07 | 3.66 | 5.23 | 14.32 | 35.40 | 41.91 | 20.43 | 35.37 | 12.32 | 16.69 |
| 50 | 3 | 2.50 | 0.49 | 1.08 | 6.12 | 27.80 | 39.44 | 8.90 | 30.97 | 4.78 | 9.19 |
| 100 | 3 | 13.90 | 8.87 | 10.91 | 16.88 | 30.11 | 29.68 | 28.97 | 30.31 | 15.13 | 15.89 |
| 100 | 3 | 7.65 | 3.98 | 5.40 | 10.40 | 23.81 | 27.38 | 19.72 | 25.59 | 8.42 | 9.38 |
| 100 | 3 | 2.00 | 0.38 | 1.07 | 3.70 | 15.31 | 24.14 | 8.15 | 19.84 | 2.35 | 3.21 |
| 200 | 3 | 11.33 | 8.34 | 9.67 | 13.25 | 21.59 | 20.11 | 28.59 | 23.10 | 13.09 | 13.07 |
| 200 | 3 | 5.77 | 3.59 | 4.41 | 7.29 | 14.81 | 17.28 | 19.65 | 17.54 | 7.27 | 7.38 |
| 200 | 3 | 1.08 | 0.29 | 0.73 | 2.02 | 6.88 | 13.37 | 8.59 | 10.91 | 1.73 | 1.91 |
| 400 | 3 | 10.97 | 8.51 | 9.71 | 12.42 | 16.82 | 12.63 | 27.73 | 17.37 | 11.76 | 11.46 |
| 400 | 3 | 5.55 | 4.04 | 4.76 | 6.86 | 10.33 | 10.09 | 19.21 | 11.44 | 6.09 | 5.70 |
| 400 | 3 | 1.27 | 0.62 | 0.93 | 2.03 | 3.41 | 6.35 | 8.31 | 5.25 | 1.34 | 1.25 |
| 50 | 5 | 15.66 | 8.60 | 10.78 | 20.06 | 46.74 | 49.84 | 42.29 | 43.25 | 20.47 | 24.45 |
| 50 | 5 | 8.84 | 3.40 | 5.26 | 13.30 | 41.10 | 48.57 | 33.72 | 39.97 | 12.78 | 17.47 |
| 50 | 5 | 2.39 | 0.60 | 0.95 | 5.39 | 33.35 | 46.88 | 19.83 | 36.37 | 4.39 | 8.84 |
| 100 | 5 | 12.57 | 7.83 | 9.69 | 15.43 | 33.58 | 32.39 | 41.09 | 33.44 | 15.68 | 16.66 |
| 100 | 5 | 6.54 | 3.34 | 4.50 | 9.08 | 27.64 | 30.48 | 32.36 | 29.05 | 9.33 | 10.31 |
| 100 | 5 | 1.47 | 0.33 | 0.78 | 2.81 | 19.58 | 27.78 | 18.26 | 23.64 | 2.63 | 3.66 |
| 200 | 5 | 11.98 | 8.40 | 9.85 | 14.04 | 23.99 | 20.07 | 41.34 | 24.45 | 12.36 | 12.45 |
| 200 | 5 | 6.23 | 3.96 | 5.07 | 8.05 | 17.54 | 17.63 | 32.78 | 19.12 | 6.70 | 6.87 |
| 200 | 5 | 1.46 | 0.52 | 0.94 | 2.42 | 9.40 | 14.54 | 18.46 | 12.77 | 1.68 | 1.76 |
| 400 | 5 | 11.65 | 9.08 | 10.24 | 13.39 | 19.06 | 14.06 | 40.36 | 19.73 | 12.15 | 11.78 |
| 400 | 5 | 5.88 | 4.02 | 4.73 | 7.30 | 12.71 | 11.88 | 31.76 | 13.97 | 6.31 | 6.04 |
| 400 | 5 | 1.33 | 0.60 | 0.85 | 2.17 | 5.07 | 9.02 | 18.36 | 7.70 | 1.45 | 1.20 |
| 50 | 7 | 15.44 | 9.03 | 10.83 | 19.94 | 46.98 | 49.16 | 49.21 | 42.94 | | |
| 50 | 7 | 8.90 | 3.82 | 5.24 | 13.18 | 41.65 | 48.28 | 41.59 | 39.55 | | |
| 50 | 7 | 2.40 | 0.75 | 1.10 | 5.26 | 33.83 | 46.96 | 27.69 | 36.21 | | |
| 100 | 7 | 13.20 | 7.92 | 10.10 | 16.14 | 33.98 | 31.14 | 50.06 | 33.09 | | |
| 100 | 7 | 6.83 | 3.50 | 4.78 | 9.68 | 27.92 | 29.41 | 41.40 | 28.45 | | |
| 100 | 7 | 1.71 | 0.39 | 0.90 | 3.58 | 20.35 | 27.38 | 26.89 | 23.59 | | |
| 200 | 7 | 11.79 | 8.47 | 9.92 | 13.96 | 25.29 | 19.27 | 50.06 | 25.44 | | |
| 200 | 7 | 6.28 | 3.37 | 4.55 | 8.24 | 18.69 | 17.63 | 42.07 | 20.16 | | |
| 200 | 7 | 1.25 | 0.41 | 0.82 | 2.15 | 11.22 | 15.28 | 26.93 | 14.10 | | |
| 400 | 7 | 11.16 | 8.44 | 9.51 | 12.95 | 18.97 | 11.97 | 49.43 | 19.40 | | |
| 400 | 7 | 5.49 | 3.90 | 4.64 | 6.88 | 12.76 | 10.13 | 40.70 | 14.08 | | |
| 400 | 7 | 1.26 | 0.59 | 0.83 | 2.06 | 5.74 | 7.96 | 25.83 | 7.84 | | |

Table 5.9: Null rejection rates of the four tests for gamma models

| | | GQ model | | | | NPML model | | | | VC model | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{T}}$ |
| 50 | 3 | 13.82 | 24.11 | 12.27 | 15.72 | 37.34 | 68.48 | 7.07 | 27.10 | 16.71 | 20.42 |
| 50 | 3 | 7.74 | 16.50 | 6.75 | 8.29 | 27.80 | 62.10 | 3.52 | 17.50 | 9.81 | 11.66 |
| 50 | 3 | 2.05 | 7.36 | 1.76 | 1.88 | 13.48 | 51.06 | 0.93 | 6.27 | 2.90 | 2.89 |
| 100 | 3 | 12.00 | 18.55 | 11.67 | 13.14 | 22.91 | 52.00 | 5.28 | 18.62 | 13.56 | 15.75 |
| 100 | 3 | 6.41 | 11.70 | 6.49 | 6.54 | 15.02 | 43.84 | 2.89 | 10.63 | 7.44 | 8.55 |
| 100 | 3 | 1.54 | 4.49 | 2.04 | 1.40 | 6.01 | 31.11 | 0.90 | 2.80 | 1.85 | 1.76 |
| 200 | 3 | 10.78 | 15.65 | 10.23 | 11.24 | 16.57 | 38.63 | 5.20 | 13.89 | 12.27 | 13.64 |
| 200 | 3 | 5.33 | 9.33 | 5.37 | 5.49 | 10.55 | 29.63 | 2.92 | 7.69 | 6.35 | 7.02 |
| 200 | 3 | 1.18 | 2.82 | 1.64 | 1.12 | 3.62 | 17.04 | 0.79 | 1.62 | 1.55 | 1.63 |
| 400 | 3 | 10.57 | 13.07 | 10.27 | 10.70 | 14.10 | 28.43 | 5.72 | 12.22 | 12.08 | 12.73 |
| 400 | 3 | 5.60 | 7.28 | 5.29 | 5.60 | 8.04 | 20.40 | 3.02 | 6.00 | 6.37 | 6.78 |
| 400 | 3 | 1.10 | 2.25 | 1.57 | 1.09 | 2.63 | 10.04 | 0.77 | 1.28 | 1.47 | 1.18 |
| 50 | 5 | 13.96 | 24.68 | 12.91 | 15.95 | 54.63 | 82.82 | 5.72 | 35.93 | 16.86 | 19.18 |
| 50 | 5 | 7.80 | 17.07 | 7.06 | 8.76 | 44.93 | 79.12 | 2.57 | 24.95 | 9.61 | 10.87 |
| 50 | 5 | 1.94 | 7.47 | 2.00 | 2.04 | 25.99 | 71.33 | 0.36 | 10.53 | 2.88 | 2.91 |
| 100 | 5 | 12.08 | 18.65 | 11.76 | 13.26 | 36.21 | 74.34 | 2.24 | 24.69 | 14.20 | 16.50 |
| 100 | 5 | 6.40 | 11.69 | 6.46 | 6.91 | 26.30 | 68.96 | 1.15 | 15.18 | 8.04 | 9.39 |
| 100 | 5 | 1.55 | 4.28 | 1.93 | 1.53 | 12.76 | 58.38 | 0.31 | 5.44 | 2.15 | 2.12 |
| 200 | 5 | 11.04 | 14.70 | 10.93 | 11.51 | 21.68 | 60.14 | 2.11 | 17.87 | 12.90 | 14.21 |
| 200 | 5 | 5.66 | 8.94 | 5.92 | 5.77 | 14.25 | 51.95 | 0.97 | 10.45 | 6.69 | 7.55 |
| 200 | 5 | 1.15 | 2.67 | 1.55 | 0.97 | 5.47 | 38.69 | 0.23 | 2.60 | 1.50 | 1.54 |
| 400 | 5 | 11.33 | 14.35 | 10.80 | 11.56 | 15.25 | 45.07 | 2.34 | 13.03 | 11.44 | 12.33 |
| 400 | 5 | 5.77 | 8.37 | 5.87 | 5.77 | 8.79 | 36.30 | 0.92 | 6.97 | 5.78 | 6.15 |
| 400 | 5 | 1.37 | 2.52 | 1.76 | 1.23 | 3.12 | 22.54 | 0.21 | 1.66 | 1.24 | 1.22 |
| 50 | 7 | 13.56 | 24.64 | 12.76 | 15.98 | 64.94 | 89.17 | 6.11 | 42.85 | | |
| 50 | 7 | 7.30 | 17.06 | 7.01 | 8.72 | 55.77 | 86.66 | 2.48 | 32.08 | | |
| 50 | 7 | 2.12 | 7.32 | 1.89 | 1.80 | 36.67 | 81.34 | 0.36 | 16.37 | | |
| 100 | 7 | 12.13 | 18.92 | 11.26 | 13.20 | 50.27 | 84.73 | 1.41 | 30.76 | | |
| 100 | 7 | 6.45 | 11.91 | 6.37 | 6.68 | 39.55 | 81.21 | 0.45 | 20.82 | | |
| 100 | 7 | 1.47 | 4.46 | 1.79 | 1.26 | 21.51 | 73.27 | 0.04 | 8.41 | | |
| 200 | 7 | 10.83 | 14.98 | 10.08 | 11.45 | 30.26 | 76.48 | 0.69 | 20.82 | | |
| 200 | 7 | 5.47 | 8.95 | 5.52 | 5.75 | 20.97 | 70.64 | 0.27 | 12.55 | | |
| 200 | 7 | 1.19 | 2.77 | 1.78 | 1.12 | 9.35 | 59.48 | 0.05 | 4.22 | | |
| 400 | 7 | 10.82 | 13.37 | 10.05 | 10.61 | 18.51 | 61.44 | 1.04 | 15.58 | | |
| 400 | 7 | 5.74 | 8.00 | 5.61 | 5.45 | 11.01 | 53.16 | 0.33 | 8.68 | | |
| 400 | 7 | 1.50 | 2.48 | 1.87 | 1.44 | 4.08 | 39.73 | 0.09 | 2.16 | | |

Table 5.10: Null rejection rates of the four tests for normal models

| | | GQ model | | | | NPML model | | | | VC model | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{T}}$ |
| 50 | 3 | 13.36 | 14.72 | 11.94 | 11.94 | 45.14 | 75.77 | 3.79 | 25.13 | 16.78 | 15.07 |
| 50 | 3 | 7.12 | 8.13 | 6.06 | 6.06 | 34.07 | 70.43 | 1.54 | 15.50 | 9.85 | 8.45 |
| 50 | 3 | 1.78 | 2.47 | 1.12 | 1.12 | 16.70 | 60.60 | 0.24 | 4.78 | 2.53 | 1.77 |
| 100 | 3 | 11.72 | 12.22 | 11.14 | 11.14 | 24.93 | 60.04 | 2.58 | 16.67 | 13.69 | 13.16 |
| 100 | 3 | 6.07 | 6.58 | 5.59 | 5.59 | 15.70 | 52.87 | 1.21 | 9.14 | 7.55 | 6.72 |
| 100 | 3 | 1.23 | 1.51 | 1.07 | 1.07 | 5.08 | 40.06 | 0.21 | 2.51 | 1.99 | 1.51 |
| 200 | 3 | 11.45 | 11.74 | 11.16 | 11.16 | 14.72 | 41.54 | 2.92 | 13.00 | 11.72 | 11.39 |
| 200 | 3 | 5.88 | 6.05 | 5.55 | 5.55 | 8.60 | 33.62 | 1.50 | 7.04 | 6.45 | 6.05 |
| 200 | 3 | 1.21 | 1.35 | 1.08 | 1.08 | 2.36 | 20.66 | 0.32 | 1.66 | 1.47 | 1.31 |
| 400 | 3 | 10.47 | 10.62 | 10.32 | 10.32 | 12.53 | 30.08 | 4.12 | 11.72 | 10.63 | 10.47 |
| 400 | 3 | 5.36 | 5.52 | 5.24 | 5.24 | 6.73 | 21.85 | 1.93 | 6.33 | 5.47 | 5.32 |
| 400 | 3 | 1.15 | 1.21 | 1.09 | 1.09 | 1.59 | 11.31 | 0.45 | 1.38 | 1.19 | 1.11 |
| 50 | 5 | 14.05 | 15.29 | 12.86 | 12.86 | 64.70 | 89.10 | 5.32 | 41.83 | 16.40 | 14.82 |
| 50 | 5 | 7.82 | 9.11 | 6.60 | 6.60 | 55.05 | 86.40 | 1.84 | 30.72 | 9.31 | 7.82 |
| 50 | 5 | 1.76 | 2.36 | 1.13 | 1.13 | 35.35 | 80.87 | 0.23 | 15.21 | 2.40 | 1.70 |
| 100 | 5 | 11.56 | 12.05 | 11.06 | 11.06 | 52.36 | 85.67 | 0.91 | 29.34 | 13.96 | 13.17 |
| 100 | 5 | 6.06 | 6.63 | 5.49 | 5.49 | 40.55 | 82.28 | 0.31 | 19.57 | 7.77 | 7.15 |
| 100 | 5 | 1.40 | 1.66 | 1.13 | 1.13 | 21.54 | 75.34 | 0.01 | 7.19 | 2.07 | 1.72 |
| 200 | 5 | 10.96 | 11.29 | 10.75 | 10.75 | 32.44 | 79.33 | 0.34 | 19.85 | 12.15 | 11.74 |
| 200 | 5 | 5.66 | 5.96 | 5.33 | 5.33 | 22.04 | 74.51 | 0.08 | 11.49 | 6.63 | 6.23 |
| 200 | 5 | 1.21 | 1.34 | 1.05 | 1.05 | 8.35 | 64.55 | 0.00 | 3.33 | 1.54 | 1.34 |
| 400 | 5 | 9.73 | 9.81 | 9.64 | 9.64 | 20.00 | 68.50 | 0.38 | 15.88 | 11.35 | 11.14 |
| 400 | 5 | 5.13 | 5.24 | 5.07 | 5.07 | 12.00 | 61.69 | 0.11 | 8.46 | 5.59 | 5.42 |
| 400 | 5 | 1.01 | 1.03 | 0.95 | 0.95 | 3.42 | 48.17 | 0.02 | 2.14 | 1.16 | 1.05 |
| 50 | 7 | 13.71 | 14.94 | 12.43 | 12.43 | 70.11 | 92.57 | 6.53 | 47.69 | | |
| 50 | 7 | 7.37 | 8.58 | 6.42 | 6.42 | 61.41 | 90.53 | 2.67 | 37.02 | | |
| 50 | 7 | 1.93 | 2.62 | 1.26 | 1.26 | 43.16 | 86.33 | 0.31 | 21.33 | | |
| 100 | 7 | 12.16 | 12.81 | 11.52 | 11.52 | 63.61 | 91.07 | 1.01 | 37.33 | | |
| 100 | 7 | 6.54 | 7.11 | 5.96 | 5.96 | 53.28 | 89.01 | 0.25 | 26.85 | | |
| 100 | 7 | 1.66 | 1.95 | 1.31 | 1.31 | 33.24 | 84.10 | 0.03 | 12.04 | | |
| 200 | 7 | 10.68 | 10.98 | 10.42 | 10.42 | 50.34 | 89.10 | 0.13 | 27.57 | | |
| 200 | 7 | 5.72 | 6.02 | 5.50 | 5.50 | 38.51 | 86.14 | 0.05 | 17.93 | | |
| 200 | 7 | 1.24 | 1.34 | 1.10 | 1.10 | 19.52 | 80.12 | 0.00 | 6.97 | | |
| 400 | 7 | 10.22 | 10.39 | 10.07 | 10.07 | 32.26 | 84.97 | 0.02 | 19.61 | | |
| 400 | 7 | 5.11 | 5.19 | 5.03 | 5.03 | 21.84 | 81.12 | 0.00 | 11.65 | | |
| 400 | 7 | 1.05 | 1.08 | 1.01 | 1.01 | 8.86 | 73.12 | 0.00 | 3.65 | | |

Table 5.11: Null rejection rates of the four tests for inverse Gaussian models

| | | GQ model | | | | NPML model | | | | VC model | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{W}}$ | $\xi_{\mathscr{R}}$ | $\xi_{\mathscr{T}}$ | $\xi_{\mathscr{LR}}$ | $\xi_{\mathscr{T}}$ |
| 50 | 3 | 14.01 | 15.45 | 12.84 | 12.77 | 38.60 | 66.27 | 8.62 | 22.77 | 16.19 | 14.82 |
| 50 | 3 | 7.59 | 8.97 | 6.23 | 6.21 | 28.17 | 60.50 | 4.56 | 14.01 | 9.37 | 7.94 |
| 50 | 3 | 1.75 | 2.56 | 1.16 | 1.19 | 12.60 | 50.32 | 1.44 | 4.22 | 2.53 | 1.55 |
| 100 | 3 | 11.99 | 12.74 | 11.49 | 11.42 | 20.00 | 40.71 | 10.38 | 14.88 | 14.12 | 13.31 |
| 100 | 3 | 6.17 | 6.82 | 5.67 | 5.67 | 11.91 | 34.28 | 6.17 | 8.06 | 7.60 | 6.94 |
| 100 | 3 | 1.43 | 1.88 | 1.22 | 1.25 | 3.55 | 24.08 | 2.25 | 1.61 | 1.74 | 1.37 |
| 200 | 3 | 10.81 | 11.16 | 10.51 | 10.60 | 13.03 | 19.32 | 14.66 | 11.77 | 12.02 | 11.61 |
| 200 | 3 | 5.70 | 6.02 | 5.49 | 5.50 | 7.20 | 14.36 | 8.98 | 6.20 | 6.41 | 5.98 |
| 200 | 3 | 1.12 | 1.34 | 1.03 | 1.03 | 1.70 | 7.66 | 3.09 | 1.29 | 1.58 | 1.41 |
| 400 | 3 | 10.76 | 11.09 | 10.74 | 10.64 | 11.06 | 8.79 | 17.99 | 10.63 | 11.53 | 11.34 |
| 400 | 3 | 5.67 | 5.81 | 5.41 | 5.45 | 5.58 | 5.04 | 11.50 | 5.30 | 5.99 | 5.83 |
| 400 | 3 | 1.18 | 1.27 | 1.17 | 1.06 | 0.96 | 1.99 | 3.81 | 0.81 | 1.25 | 1.13 |
| 50 | 5 | 14.12 | 15.36 | 12.94 | 12.98 | 63.93 | 87.77 | 5.46 | 37.53 | 16.69 | 14.97 |
| 50 | 5 | 7.76 | 9.09 | 6.53 | 6.36 | 53.66 | 84.62 | 2.29 | 26.58 | 9.57 | 7.97 |
| 50 | 5 | 1.78 | 2.54 | 1.29 | 1.26 | 34.21 | 78.20 | 0.32 | 11.72 | 2.64 | 1.53 |
| 100 | 5 | 12.05 | 12.88 | 11.49 | 11.26 | 45.59 | 79.09 | 3.17 | 25.35 | 13.96 | 13.14 |
| 100 | 5 | 6.33 | 6.98 | 5.85 | 5.72 | 34.92 | 74.77 | 1.63 | 15.94 | 7.84 | 6.99 |
| 100 | 5 | 1.68 | 1.92 | 1.29 | 1.37 | 17.18 | 66.12 | 0.63 | 5.39 | 2.13 | 1.74 |
| 200 | 5 | 10.52 | 10.89 | 10.32 | 10.37 | 22.86 | 51.58 | 8.84 | 16.22 | 11.59 | 11.02 |
| 200 | 5 | 5.48 | 5.76 | 5.16 | 5.13 | 14.66 | 46.03 | 5.95 | 9.04 | 6.02 | 5.64 |
| 200 | 5 | 1.24 | 1.47 | 1.22 | 1.12 | 5.23 | 37.47 | 2.71 | 2.45 | 1.45 | 1.20 |
| 400 | 5 | 10.74 | 11.01 | 10.60 | 10.59 | 12.25 | 17.10 | 21.40 | 11.02 | 11.62 | 11.34 |
| 400 | 5 | 5.70 | 5.92 | 5.60 | 5.63 | 6.75 | 13.74 | 14.87 | 5.60 | 6.01 | 5.71 |
| 400 | 5 | 1.23 | 1.43 | 1.20 | 1.12 | 1.89 | 9.00 | 6.43 | 1.05 | 1.20 | 1.06 |
| 50 | 7 | 13.60 | 15.13 | 12.60 | 12.48 | 72.32 | 91.99 | 6.91 | 45.46 | | |
| 50 | 7 | 7.55 | 8.95 | 6.50 | 6.46 | 63.51 | 89.53 | 2.93 | 34.27 | | |
| 50 | 7 | 1.67 | 2.38 | 1.12 | 1.10 | 44.95 | 84.93 | 0.44 | 18.33 | | |
| 100 | 7 | 11.71 | 12.51 | 10.96 | 11.12 | 60.60 | 88.66 | 2.07 | 33.82 | | |
| 100 | 7 | 6.05 | 6.78 | 5.49 | 5.60 | 50.42 | 85.87 | 0.84 | 23.22 | | |
| 100 | 7 | 1.40 | 1.71 | 1.10 | 1.09 | 30.76 | 79.13 | 0.21 | 9.81 | | |
| 200 | 7 | 10.59 | 10.87 | 10.23 | 10.31 | 40.70 | 78.92 | 3.13 | 22.98 | | |
| 200 | 7 | 5.32 | 5.68 | 5.07 | 5.12 | 30.20 | 74.90 | 2.18 | 14.07 | | |
| 200 | 7 | 1.10 | 1.32 | 1.03 | 0.94 | 14.63 | 66.50 | 1.05 | 4.28 | | |
| 400 | 7 | 10.44 | 10.68 | 10.34 | 10.19 | 20.17 | 41.67 | 14.93 | 14.09 | | |
| 400 | 7 | 5.13 | 5.40 | 5.08 | 5.09 | 13.12 | 37.78 | 10.93 | 7.59 | | |
| 400 | 7 | 1.09 | 1.12 | 1.00 | 0.99 | 4.83 | 31.21 | 5.81 | 1.74 | | |

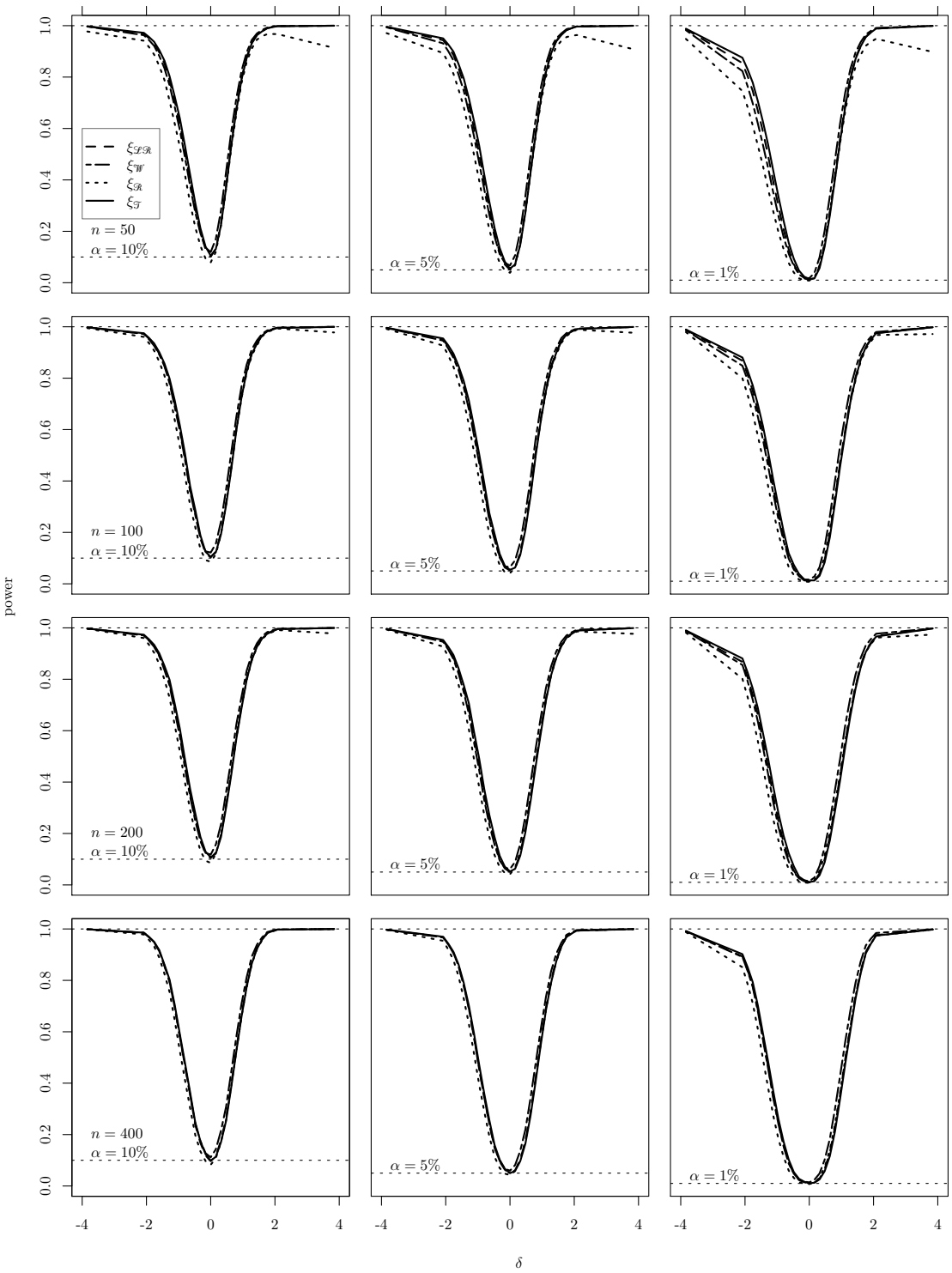


Figure 5.1: Non-null rejection rates of the four tests for Poisson response model with Gaussian quadrature fitting and $K = 3$

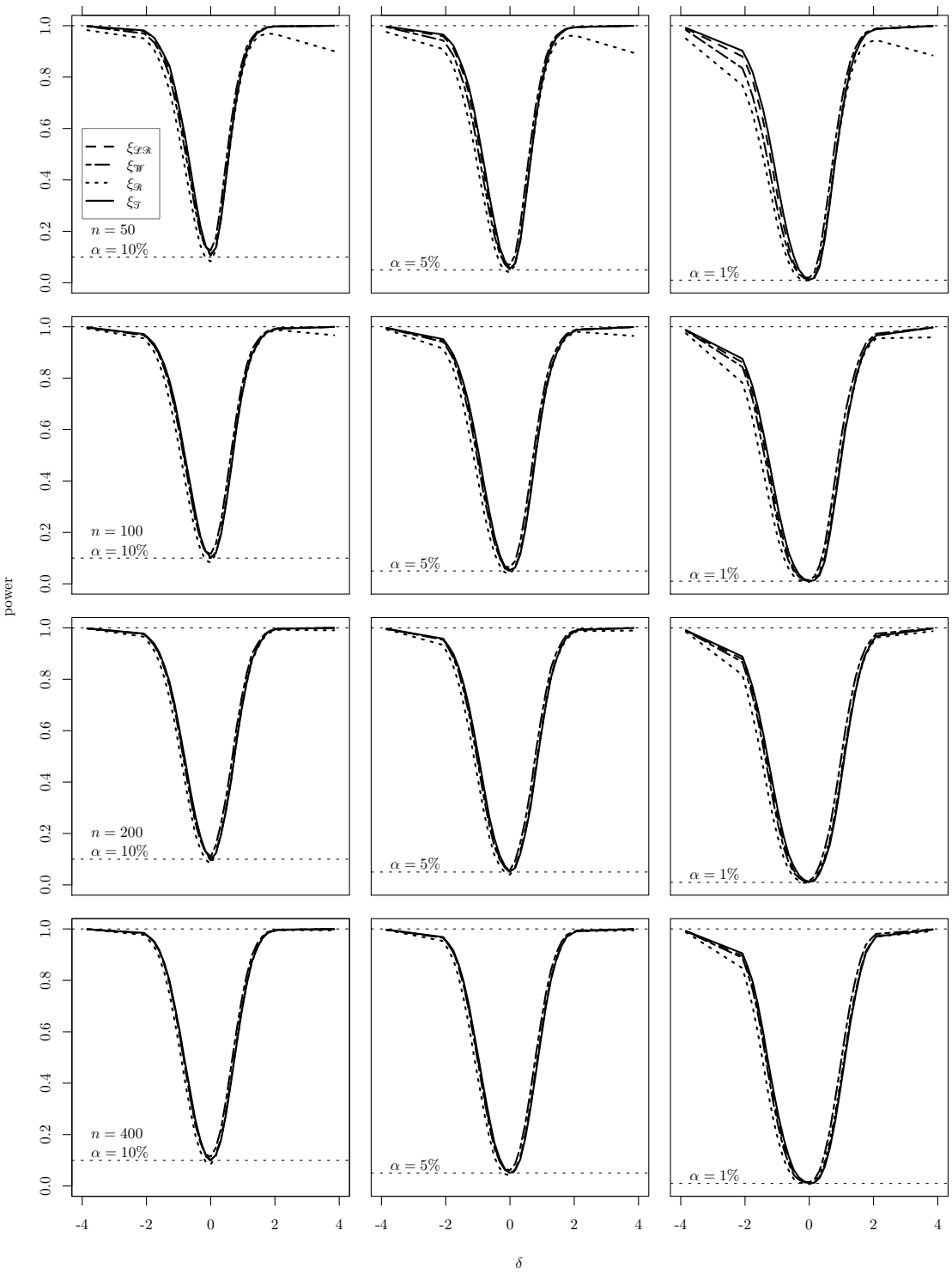


Figure 5.2: Non-null rejection rates of the four tests for Poisson response model with Gaussian quadrature fitting and $K = 5$

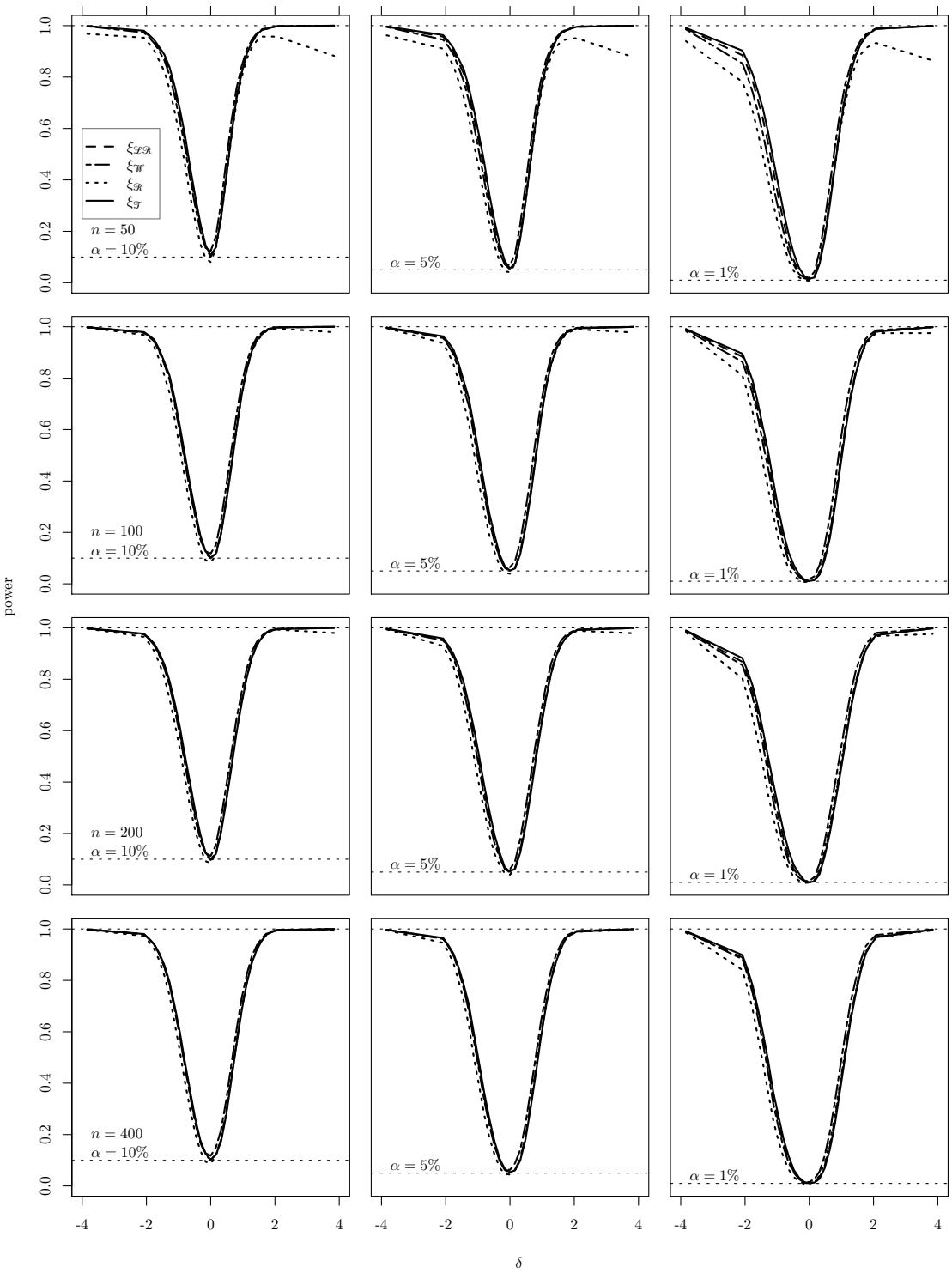


Figure 5.3: Non-null rejection rates of the four tests for Poisson response model with Gaussian quadrature fitting and $K = 7$

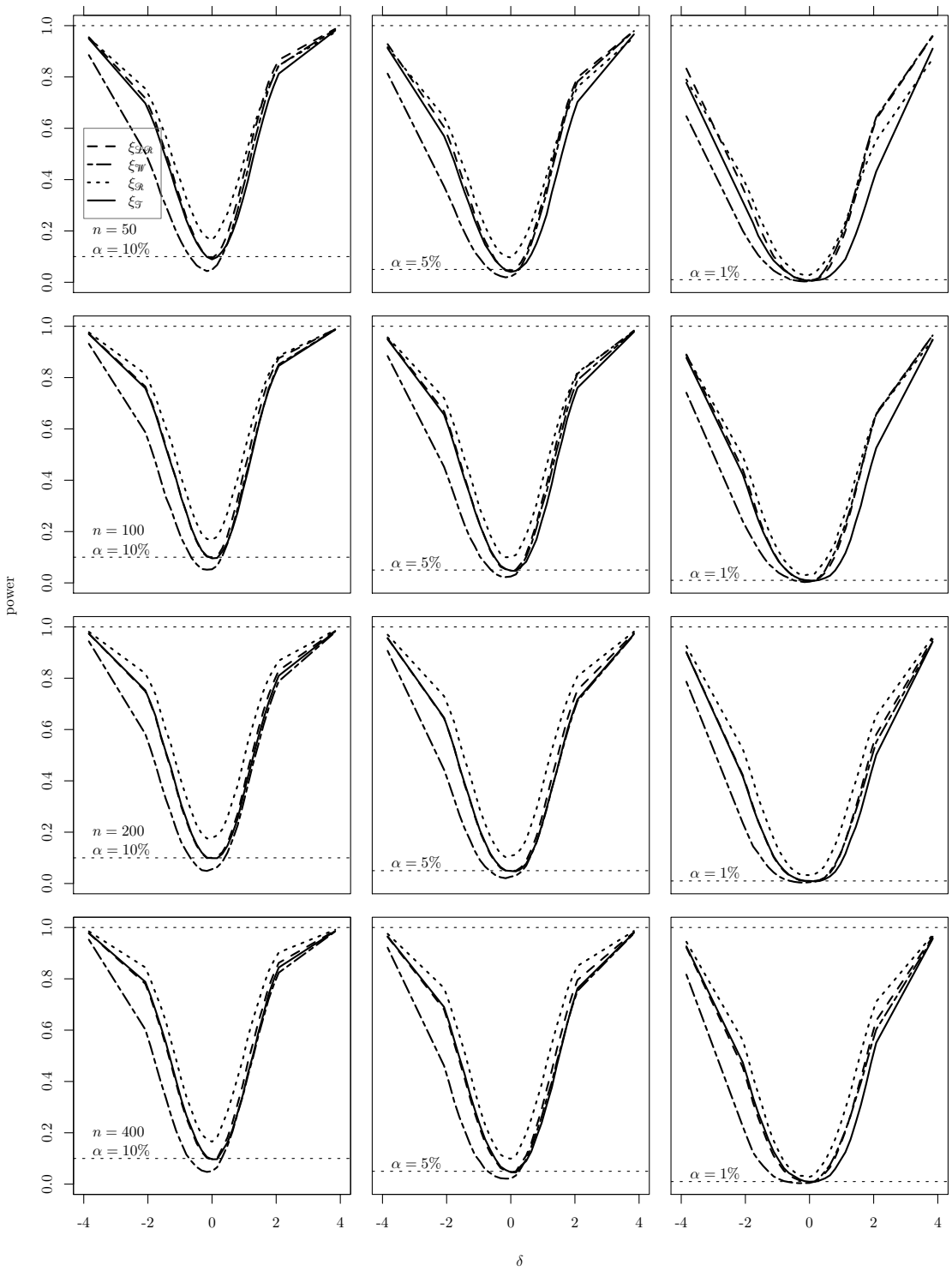


Figure 5.4: Non-null rejection rates of the four tests for Poisson response model with NPML fitting and $K = 3$

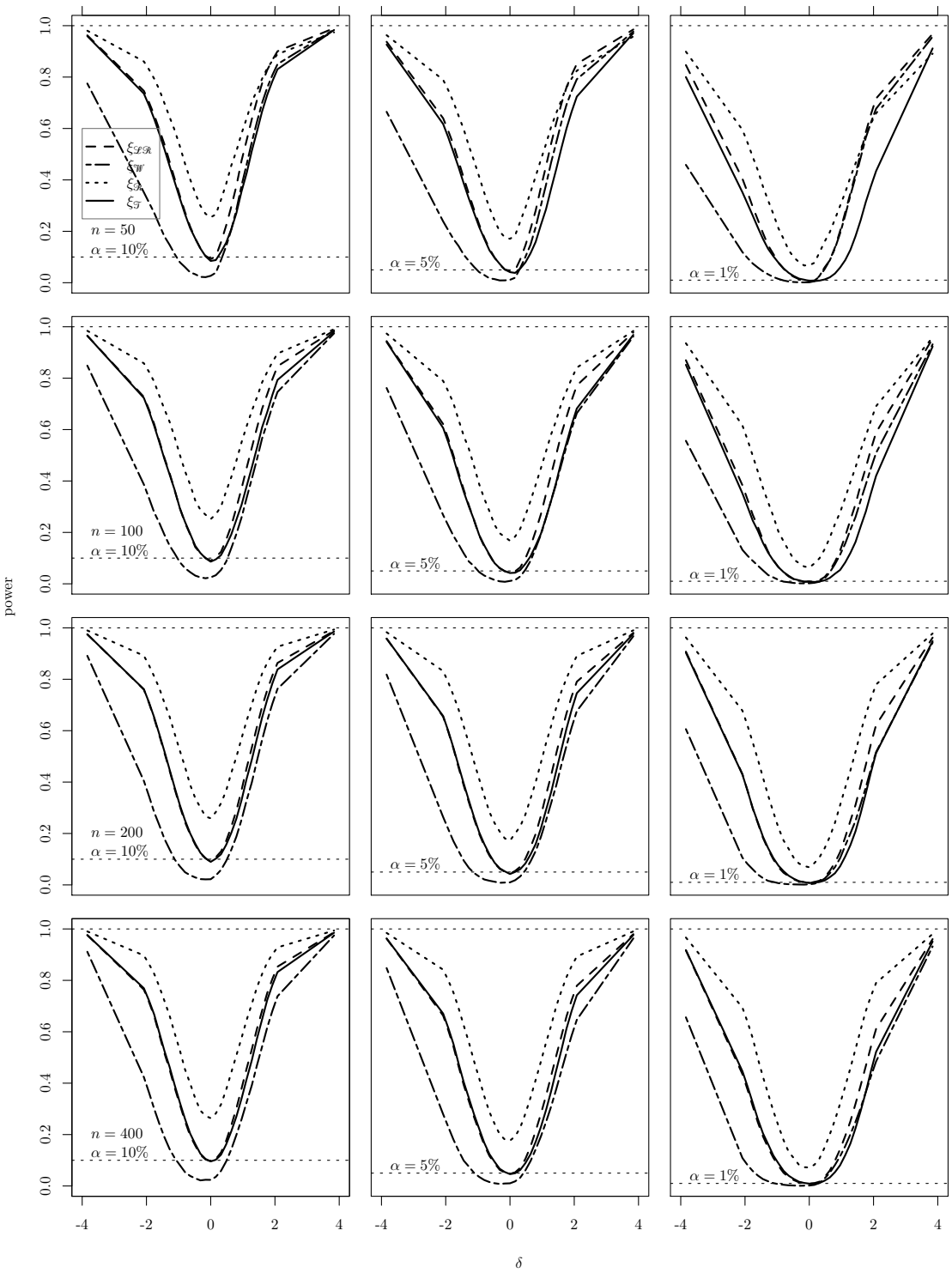


Figure 5.5: Non-null rejection rates of the four tests for Poisson response model with NPML fitting and $K = 5$

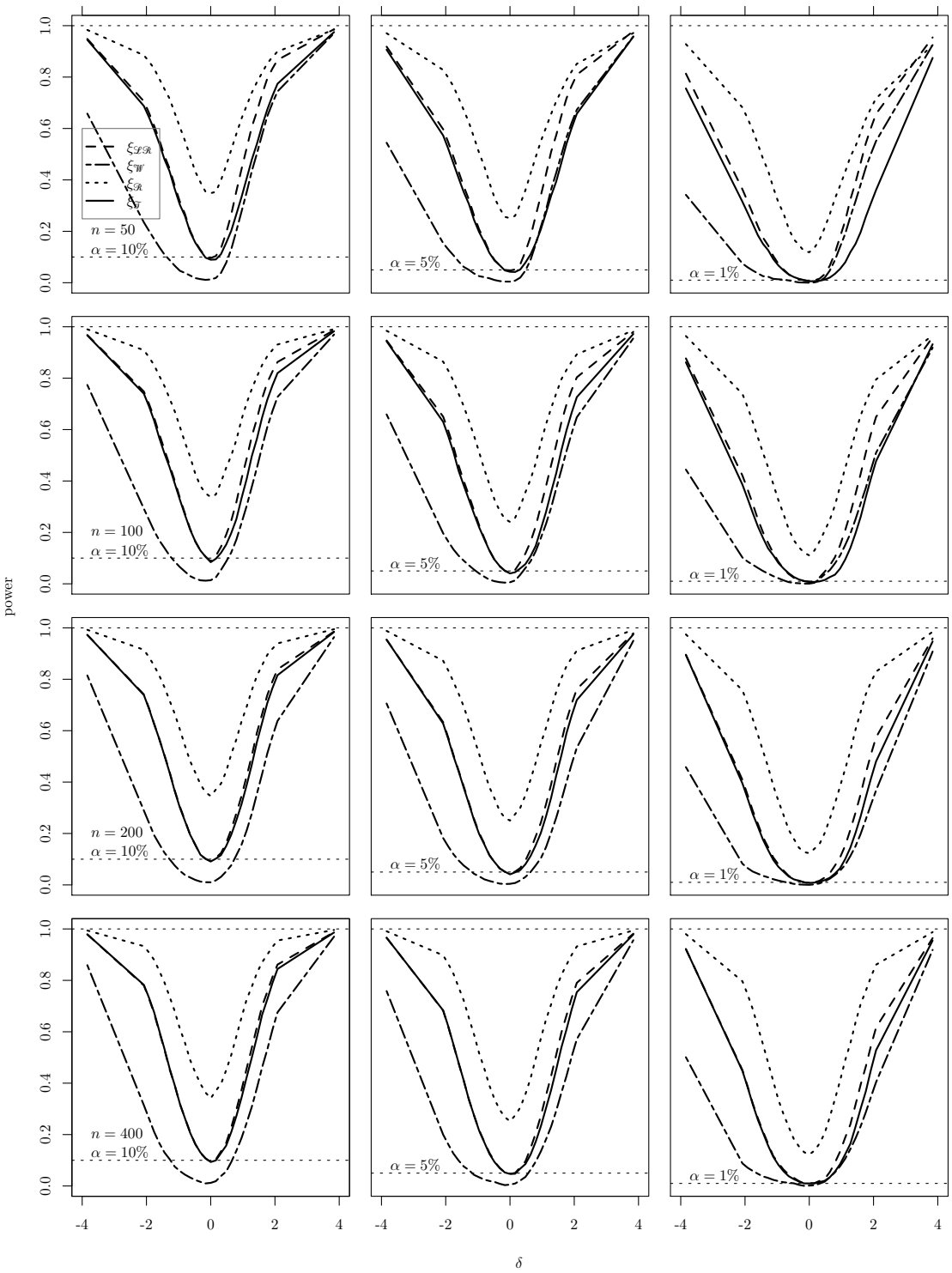


Figure 5.6: Non-null rejection rates of the four tests for Poisson response model with NPML fitting and $K = 7$

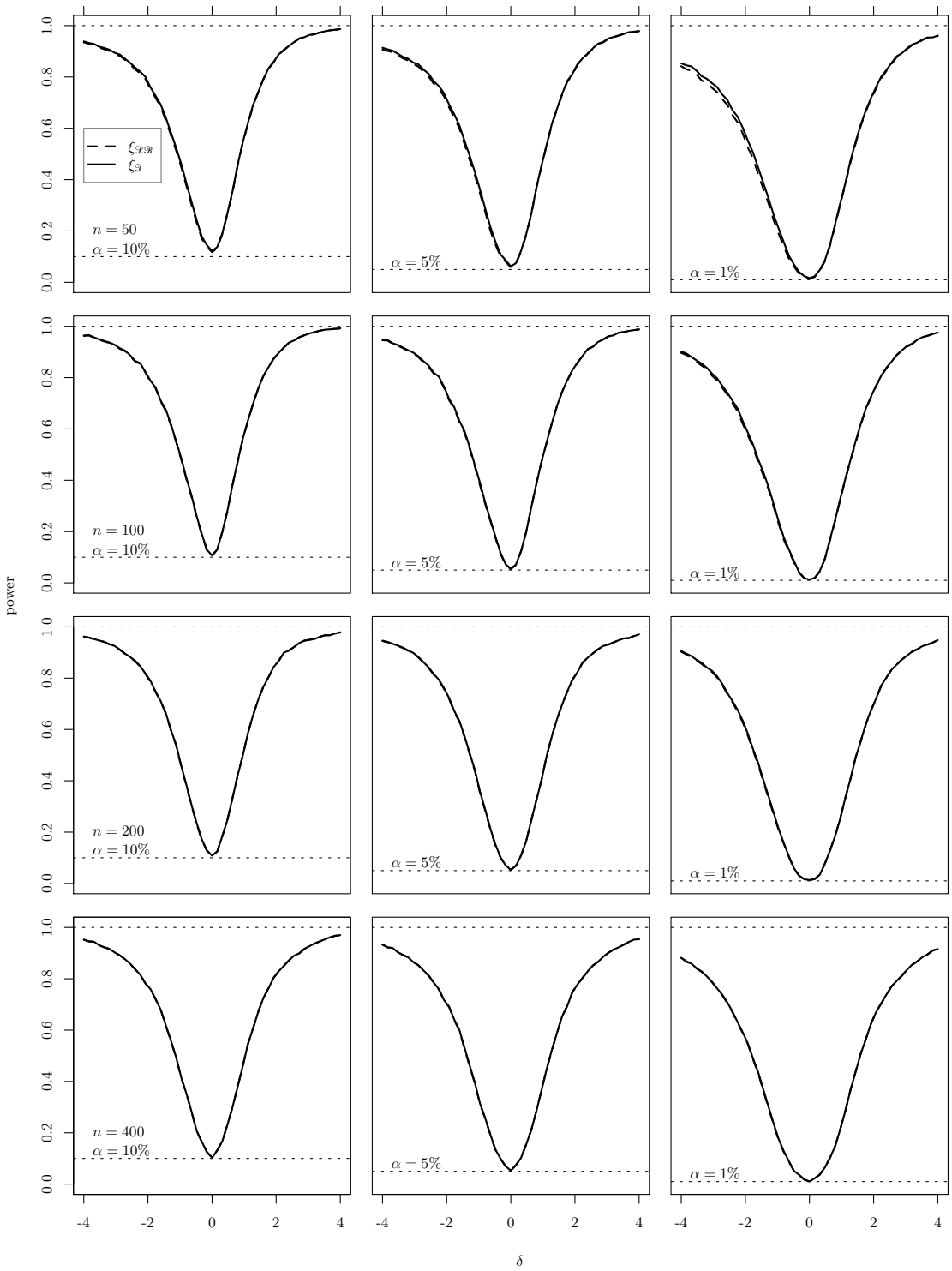


Figure 5.7: Non-null rejection rates of the four tests for Poisson response variance components model with NPML fitting and $K = 3$

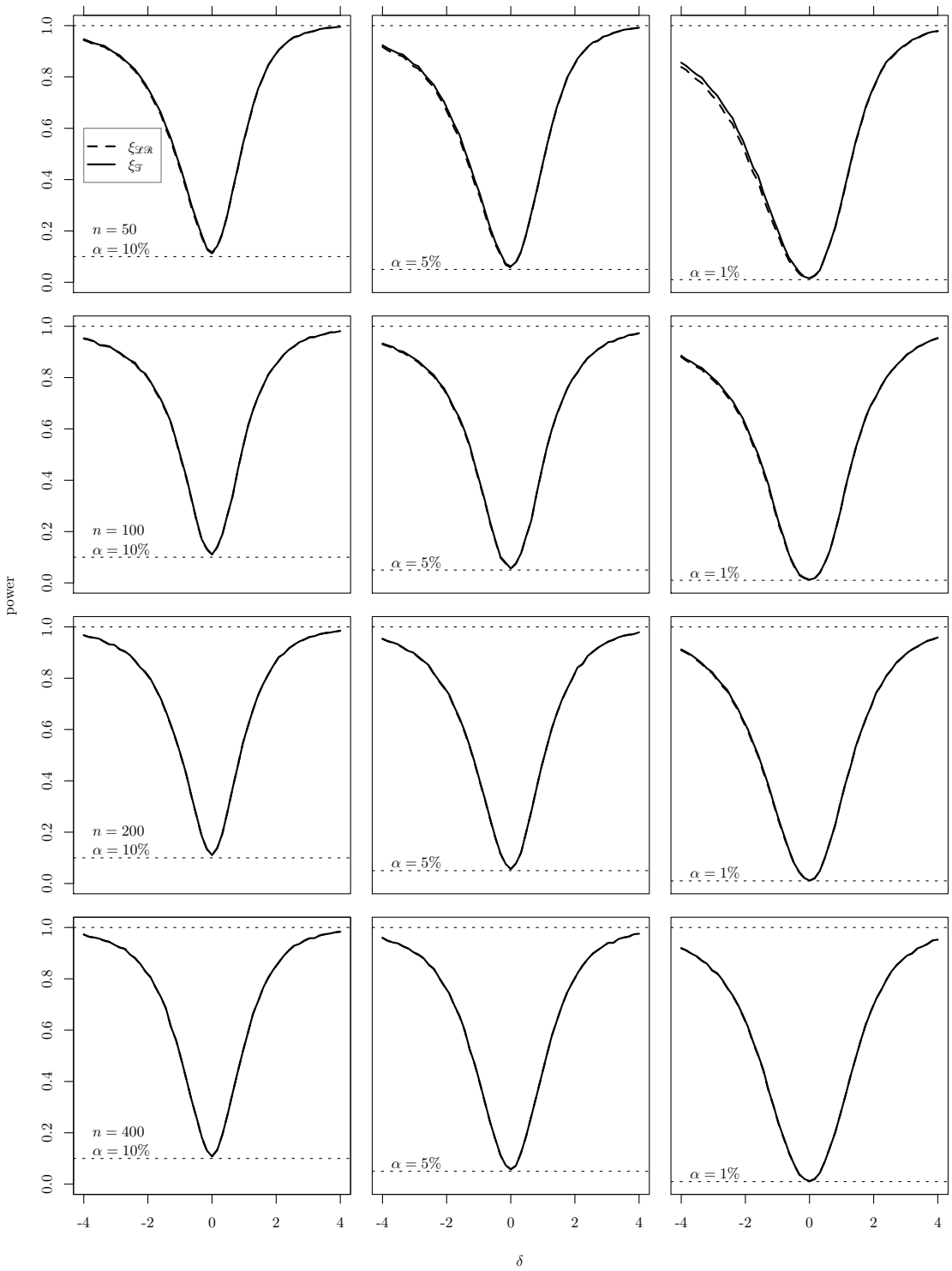


Figure 5.8: Non-null rejection rates of the four tests for Poisson response variance components model with NPML fitting and $K = 5$

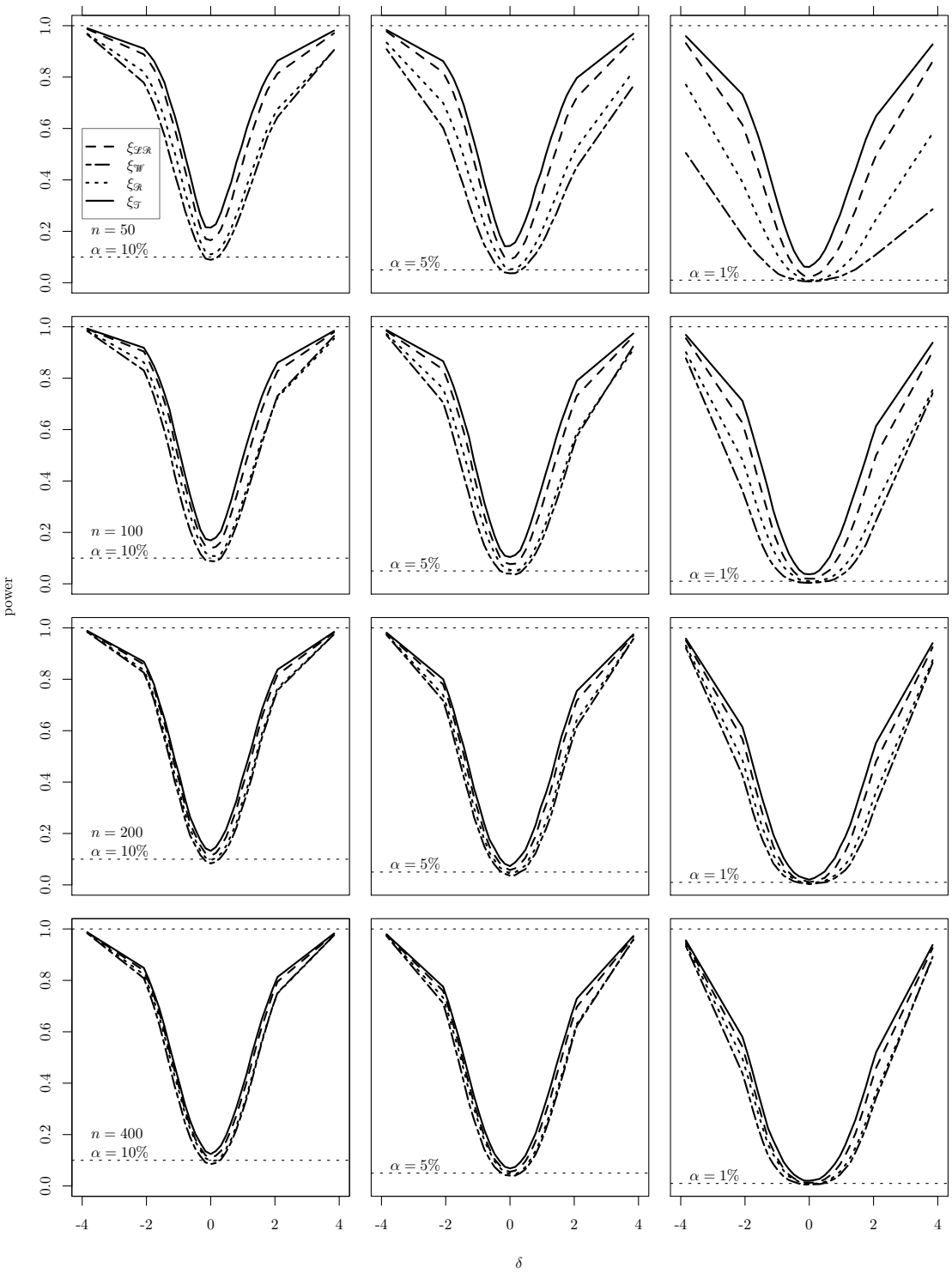


Figure 5.9: Non-null rejection rates of the four tests for binomial response model with Gaussian quadrature fitting and $K = 3$

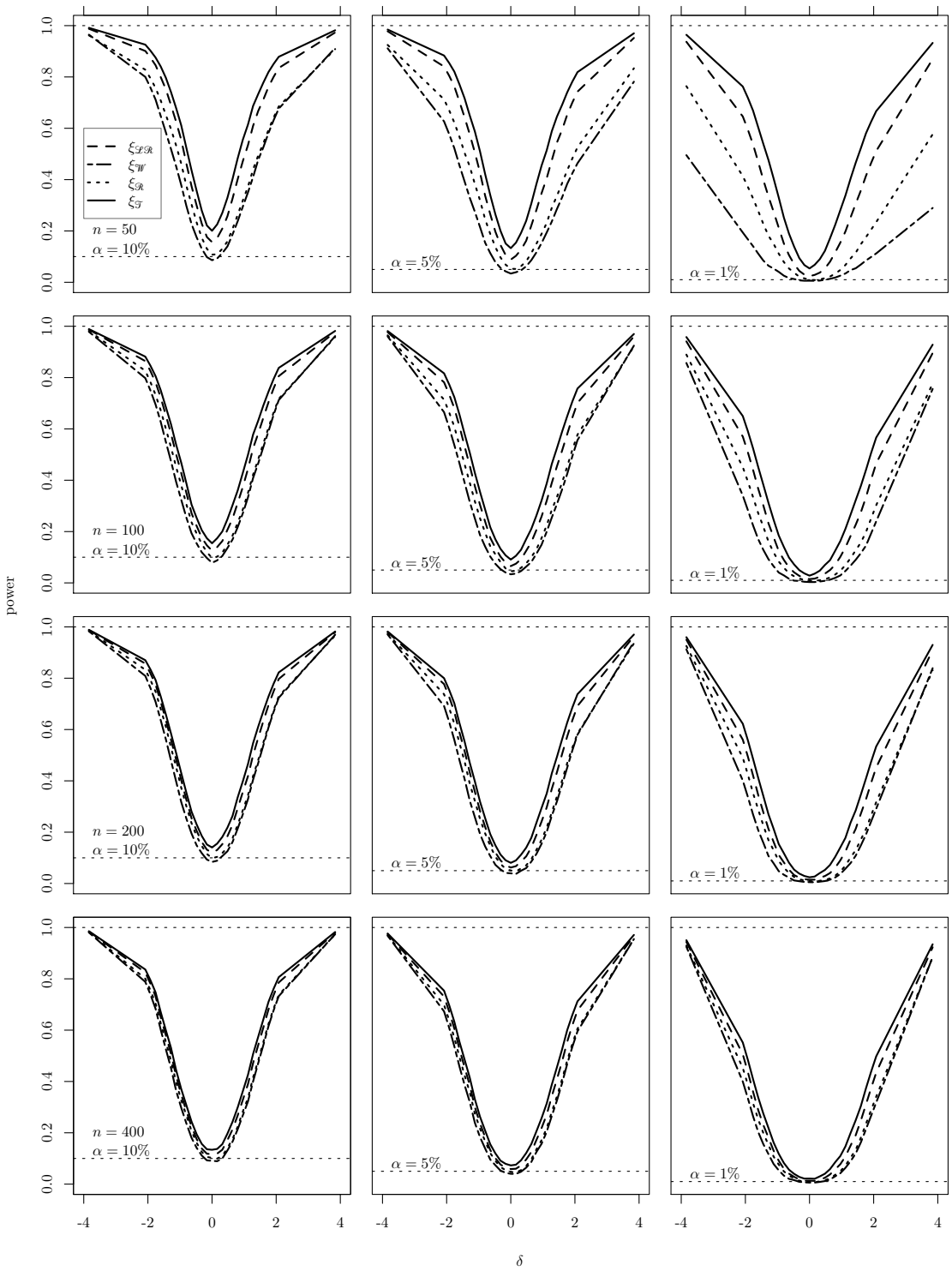


Figure 5.10: Non-null rejection rates of the four tests for binomial response model with Gaussian quadrature fitting and $K = 5$

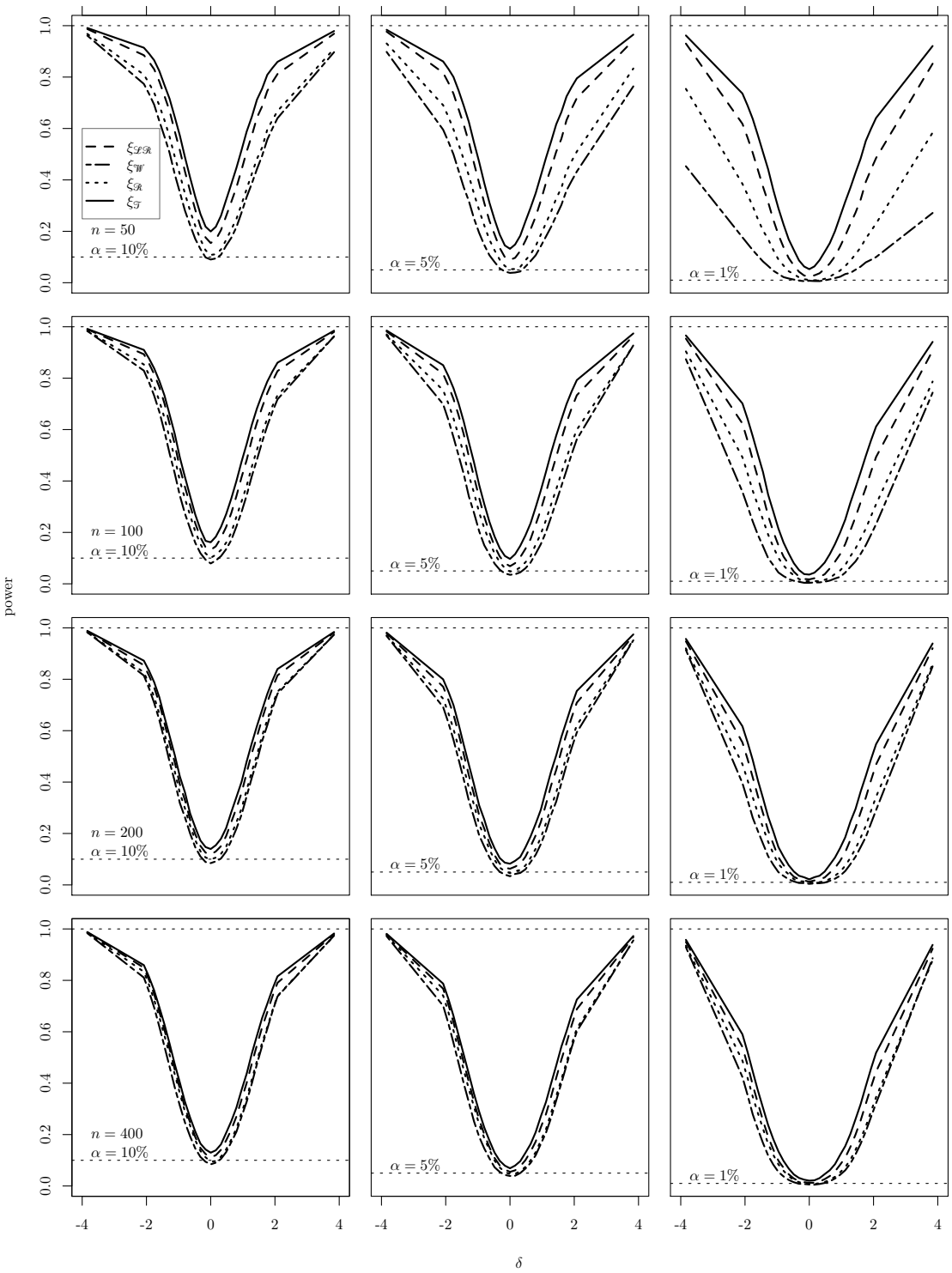


Figure 5.11: Non-null rejection rates of the four tests for binomial response model with Gaussian quadrature fitting and $K = 7$

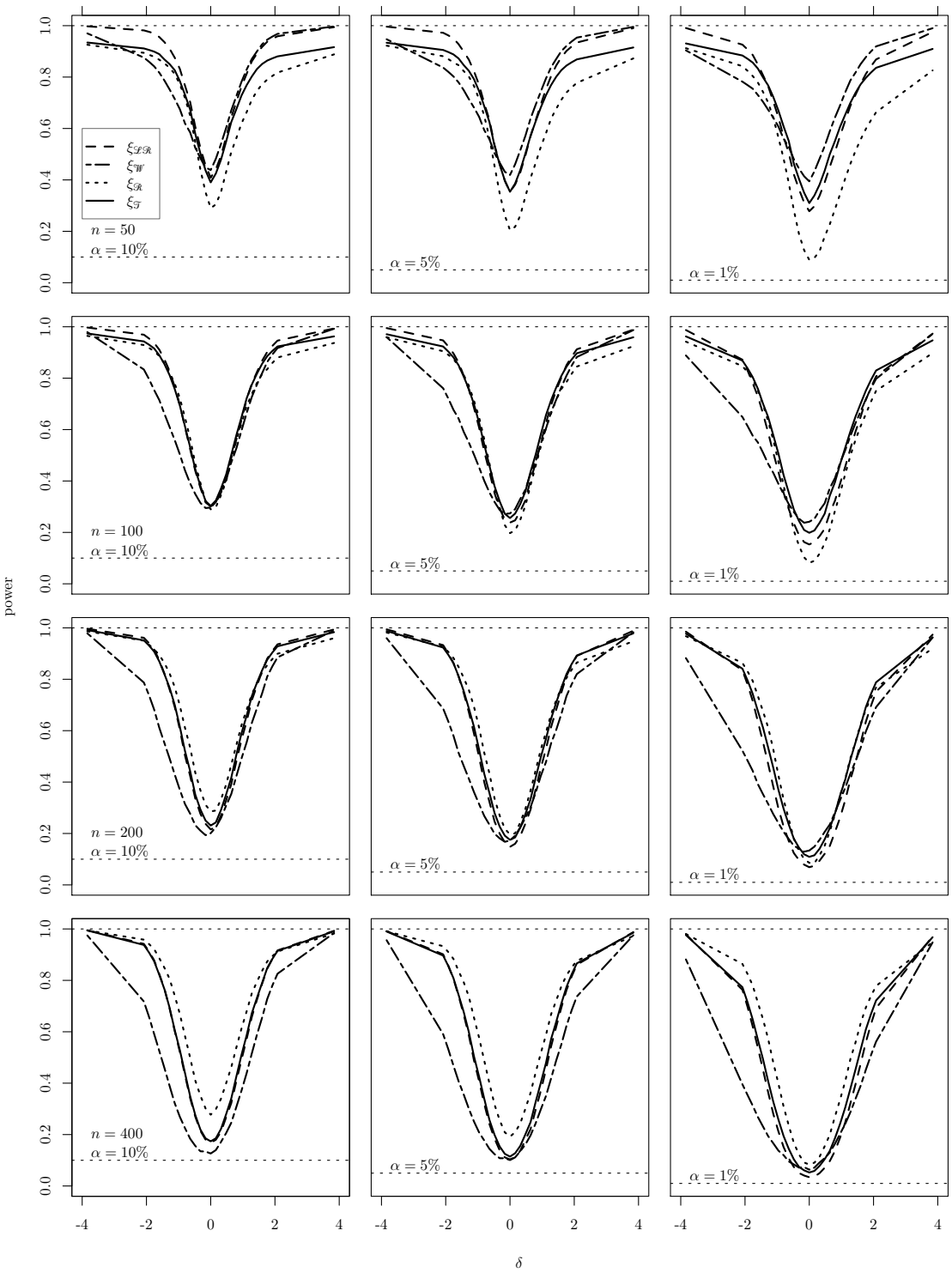


Figure 5.12: Non-null rejection rates of the four tests for binomial response model with NPML fitting and $K = 3$

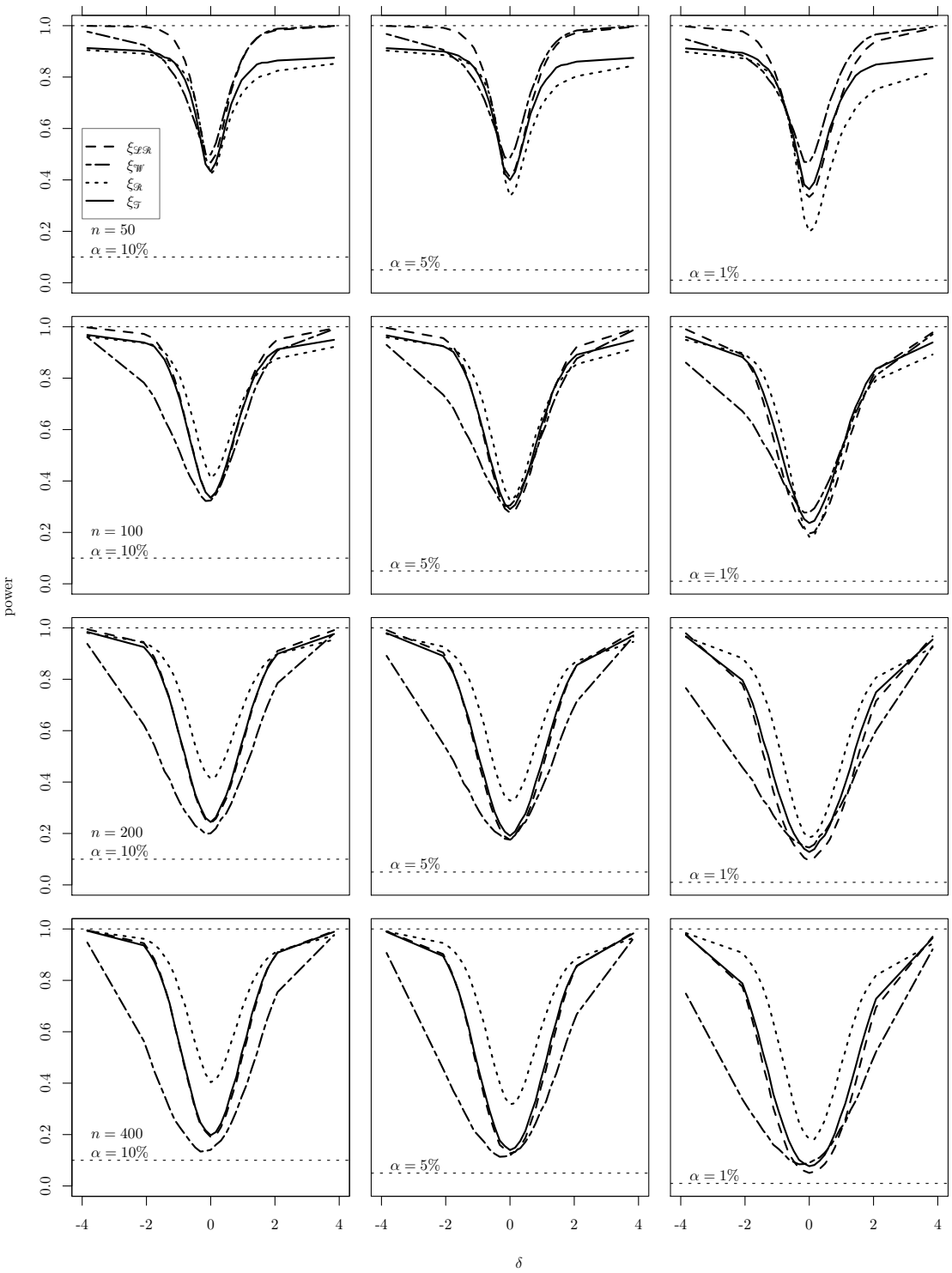


Figure 5.13: Non-null rejection rates of the four tests for binomial response model with NPML fitting and $K = 5$

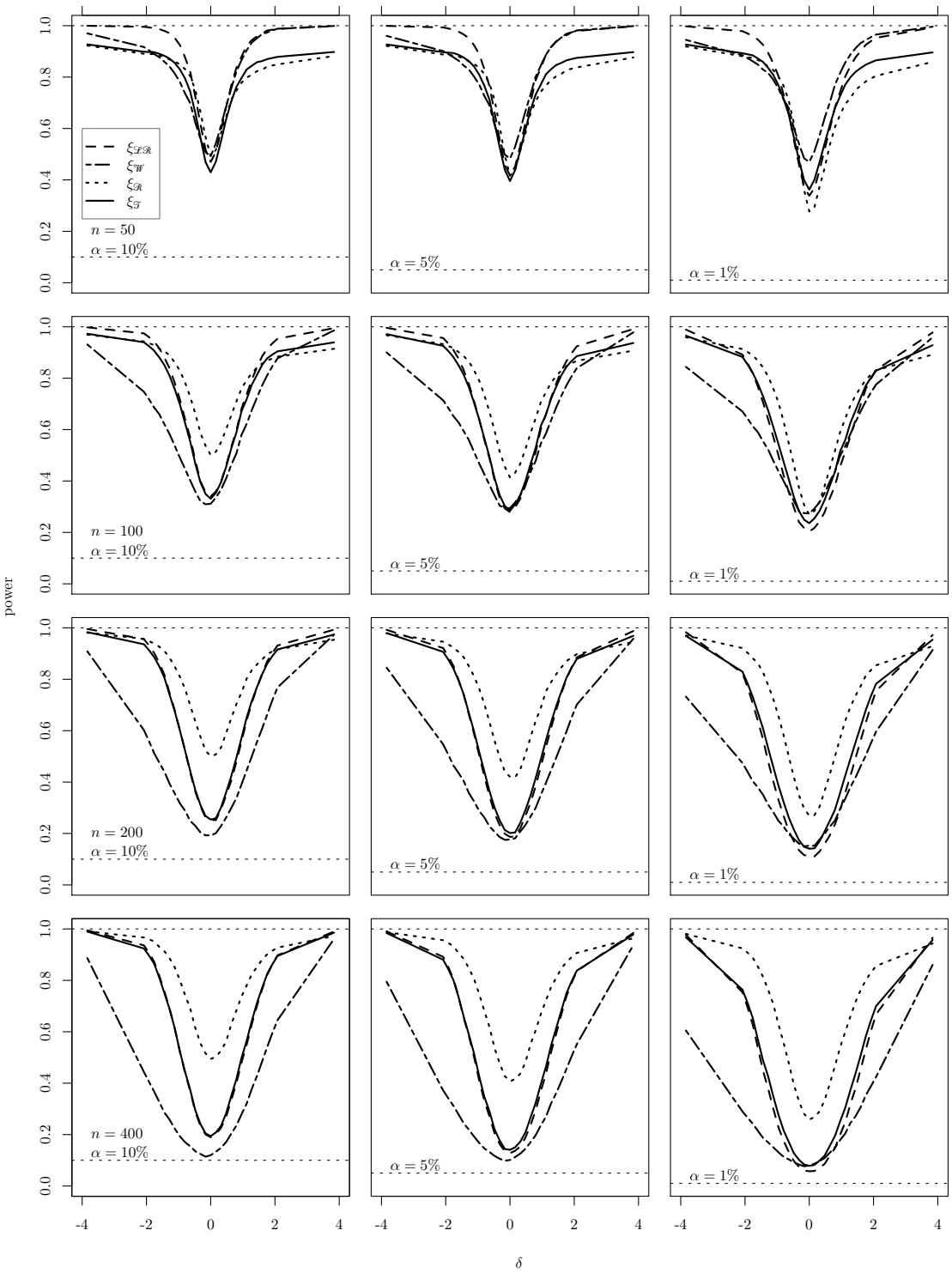


Figure 5.14: Non-null rejection rates of the four tests for binomial response model with NPML fitting and $K = 7$

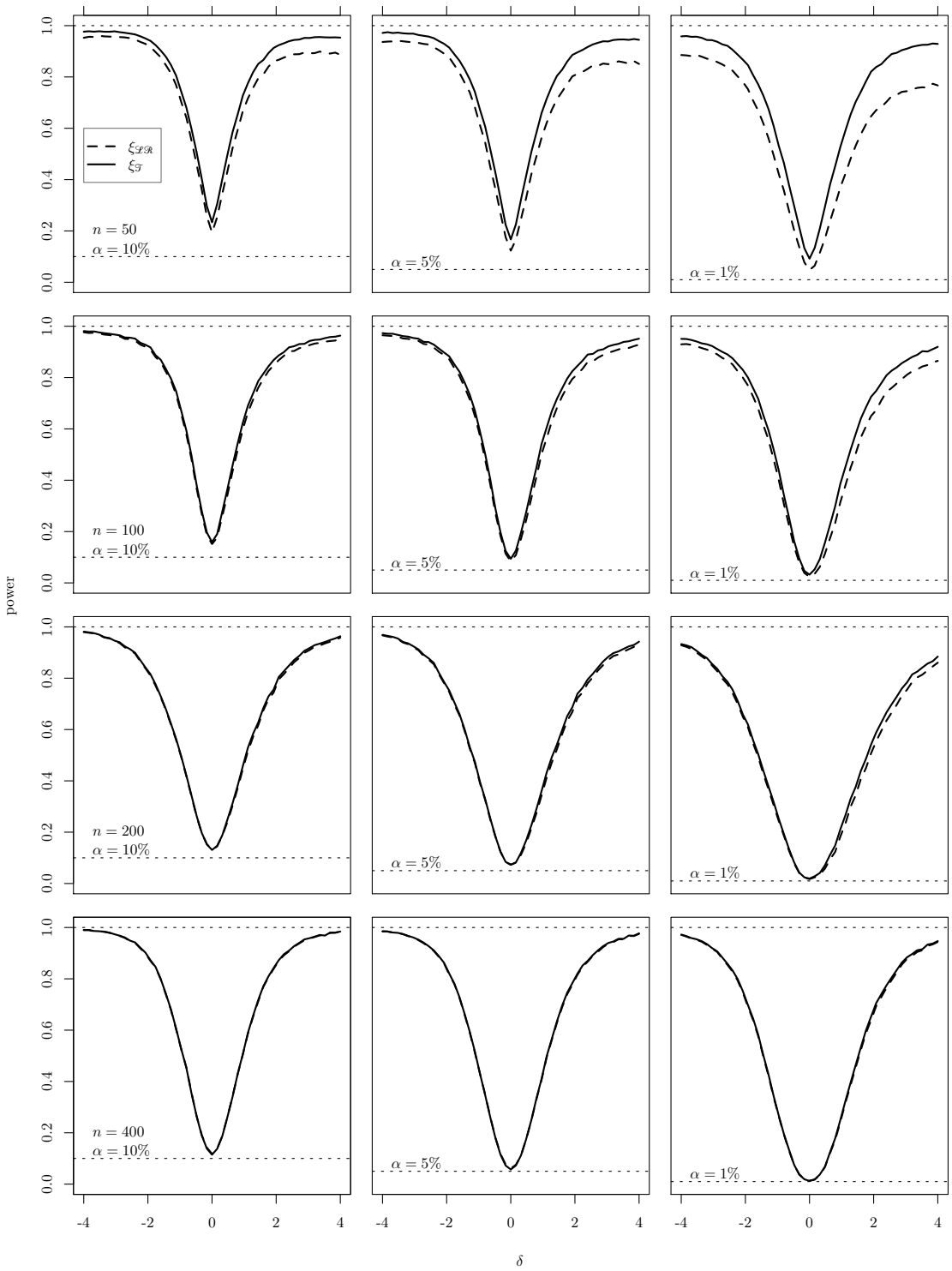


Figure 5.15: Non-null rejection rates of the four tests for binomial response variance component model with NPML fitting and $K = 3$

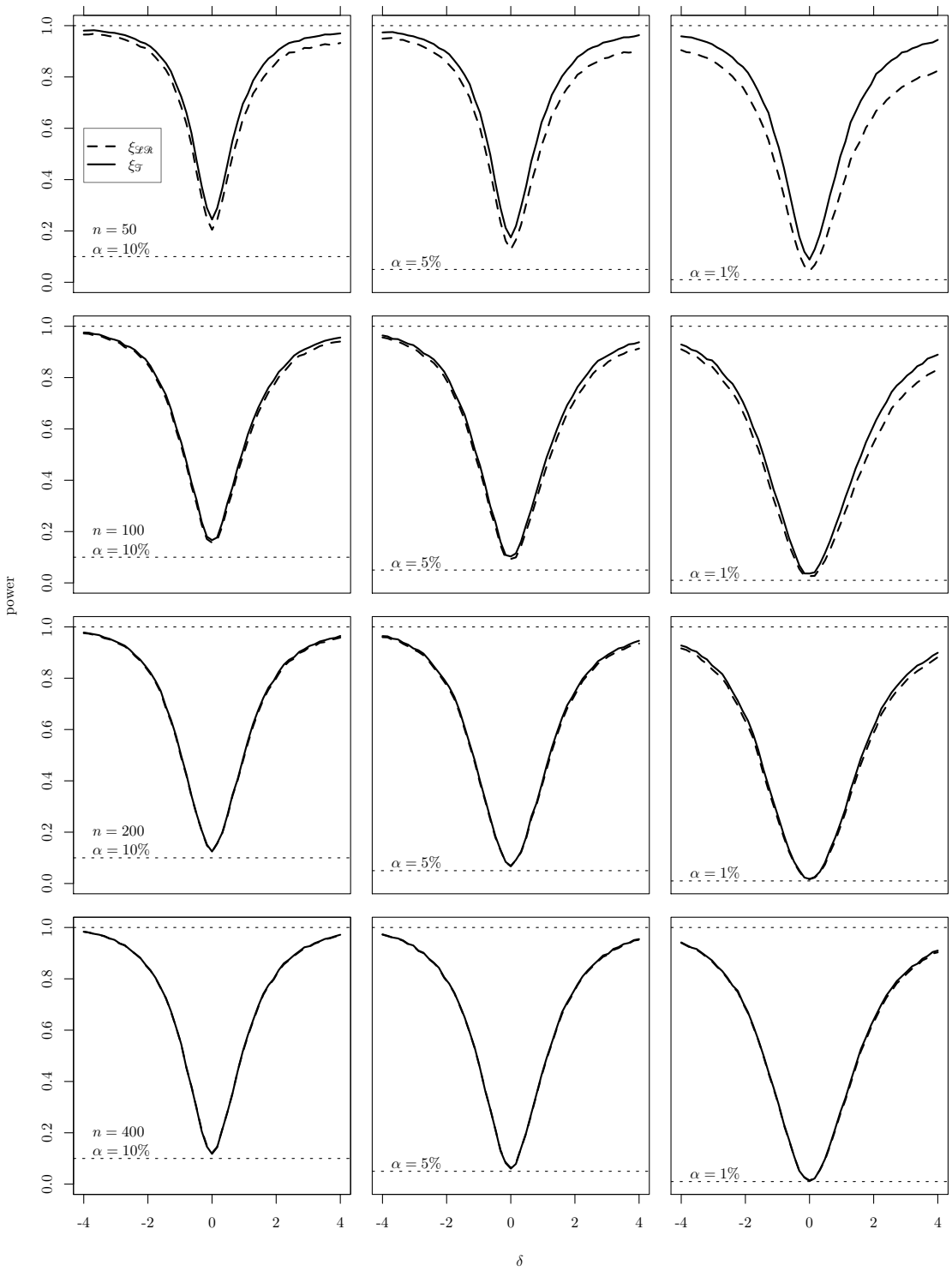


Figure 5.16: Non-null rejection rates of the four tests for binomial response variance component model with NPML fitting and $K = 5$

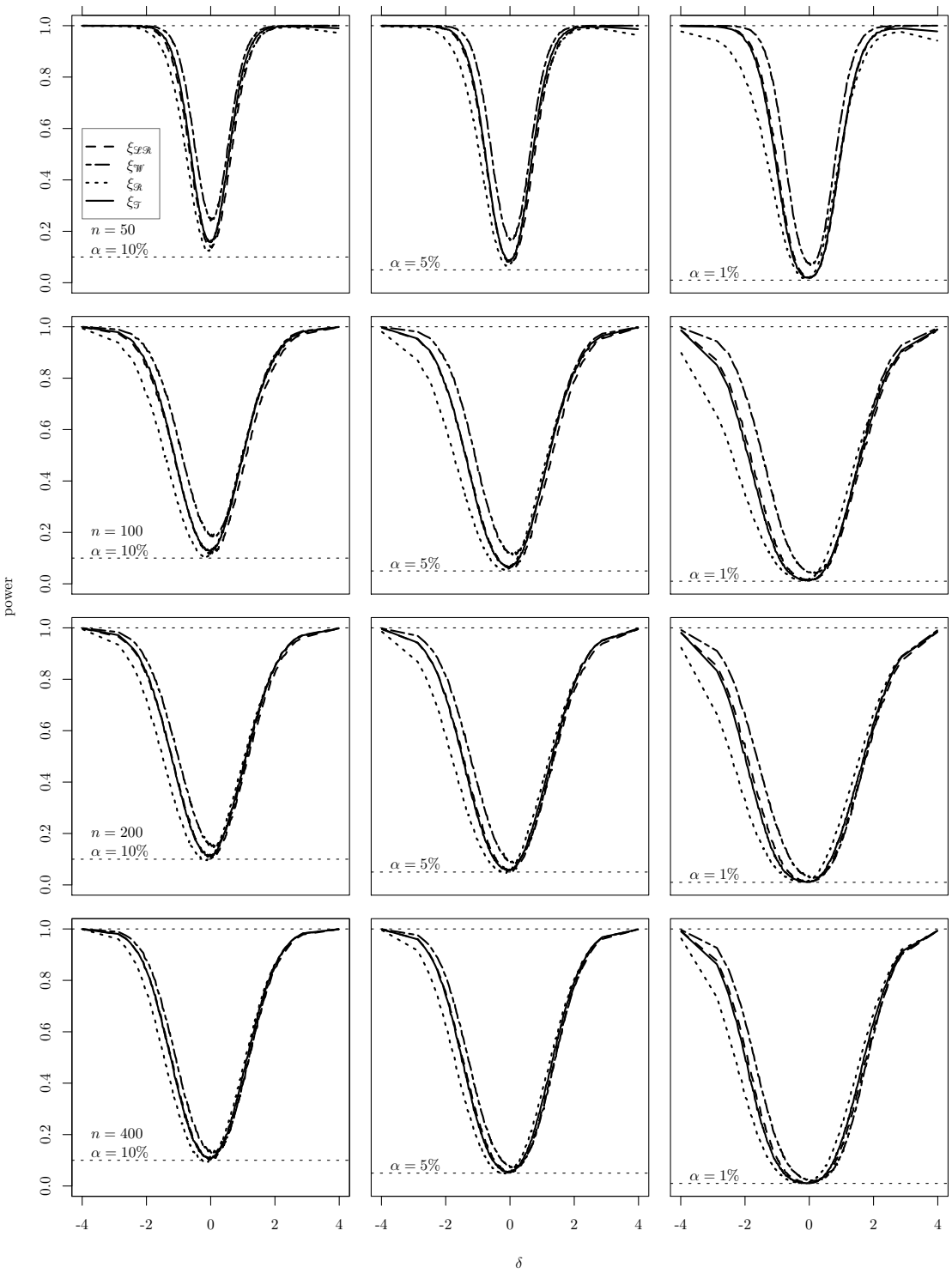


Figure 5.17: Non-null rejection rates of the four tests for gamma response model with Gaussian quadrature fitting and $K = 3$

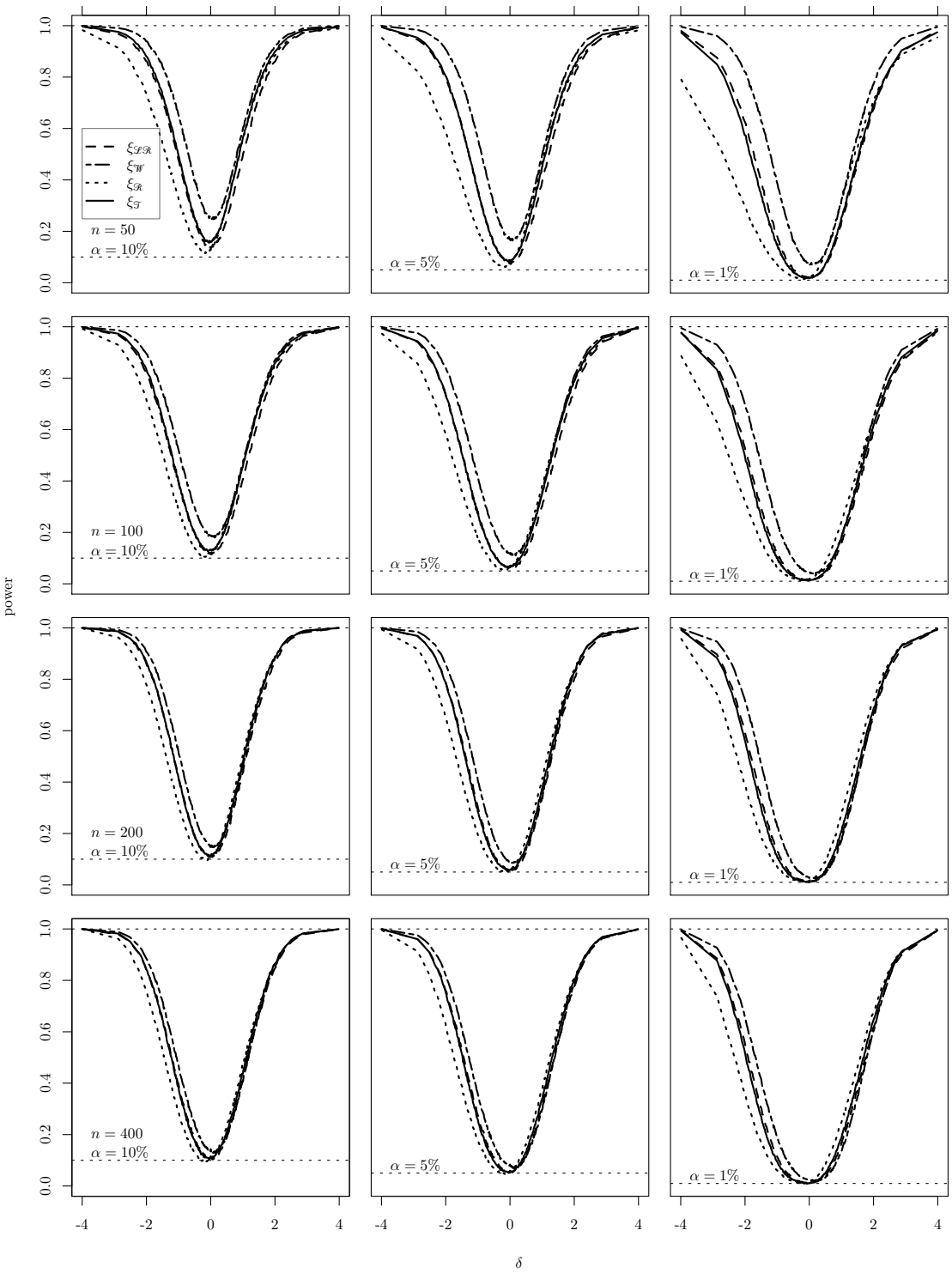


Figure 5.18: Non-null rejection rates of the four tests for gamma response model with Gaussian quadrature fitting and $K = 5$

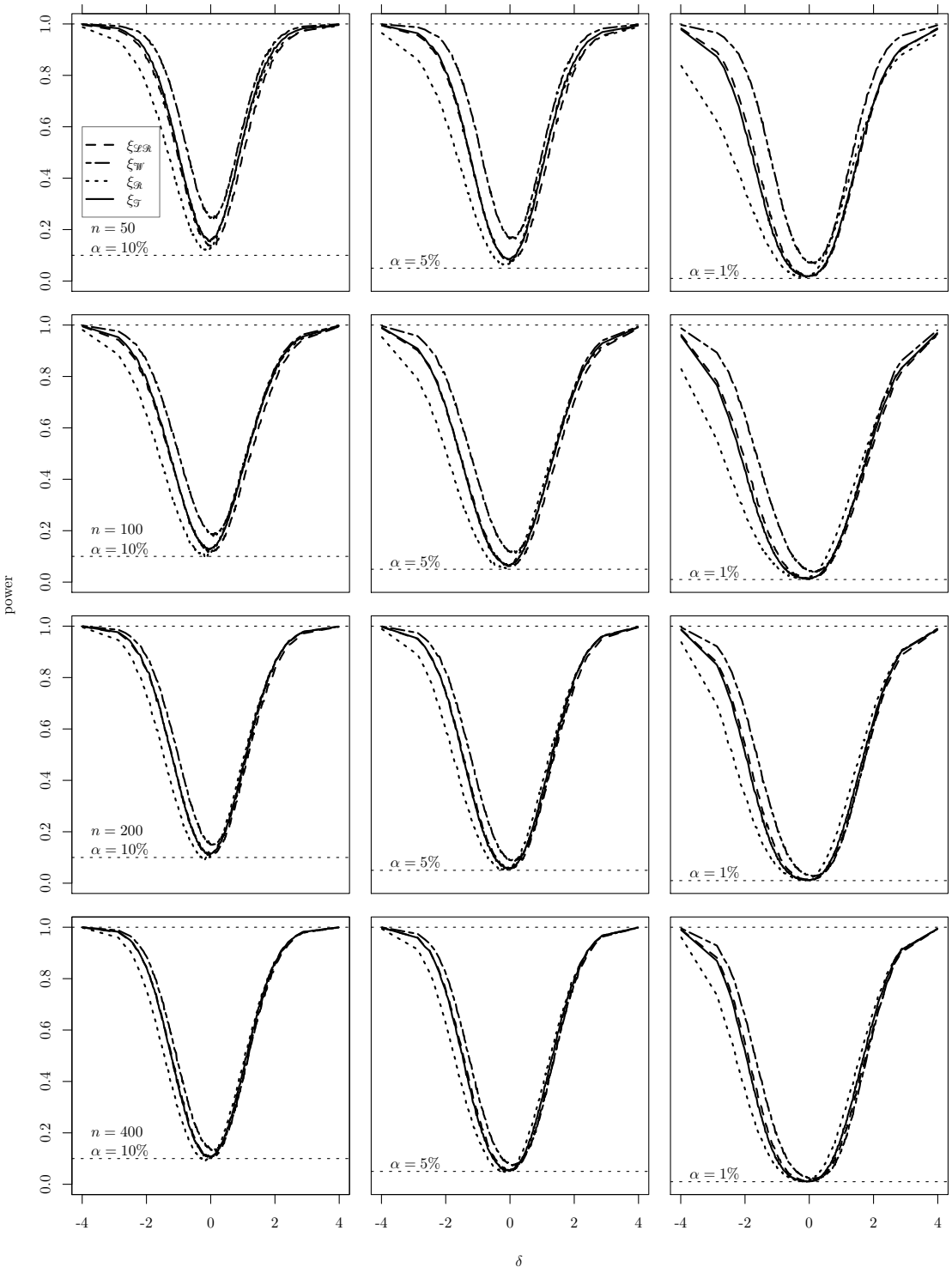


Figure 5.19: Non-null rejection rates of the four tests for gamma response model with Gaussian quadrature fitting and $K = 7$

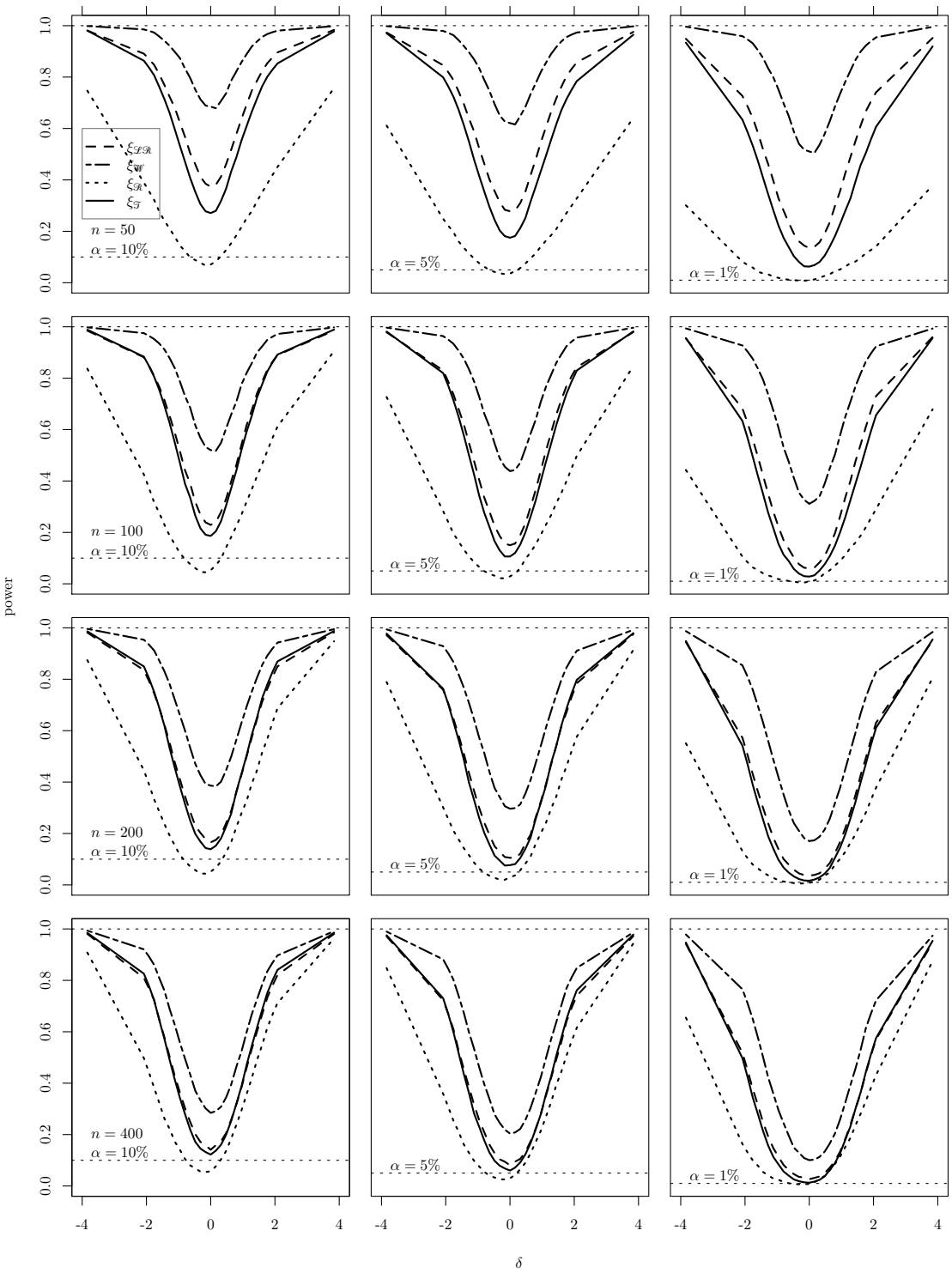


Figure 5.20: Non-null rejection rates of the four tests for gamma response model with NPML fitting and $K = 3$

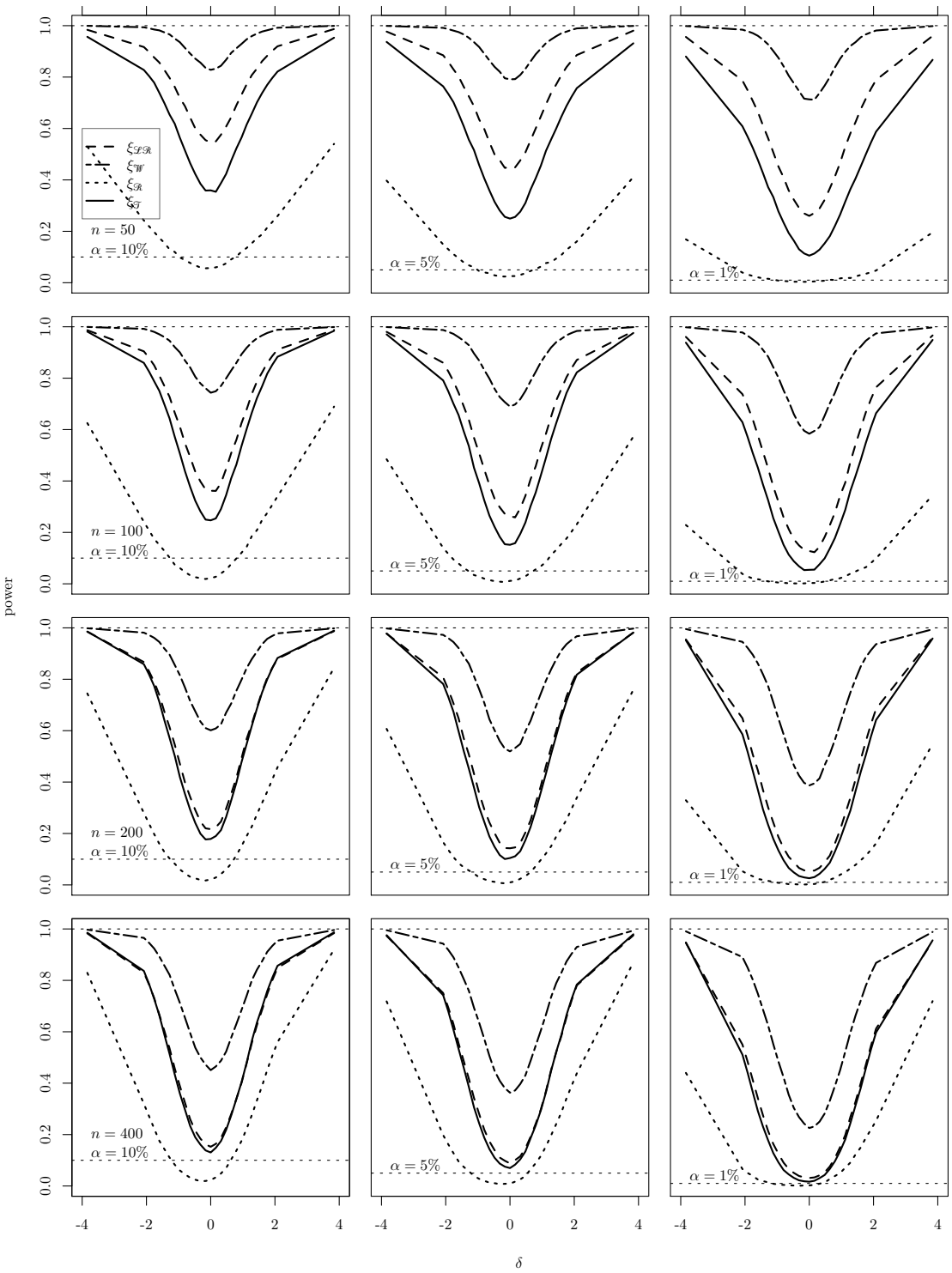


Figure 5.21: Non-null rejection rates of the four tests for gamma response model with NPML fitting and $K = 5$

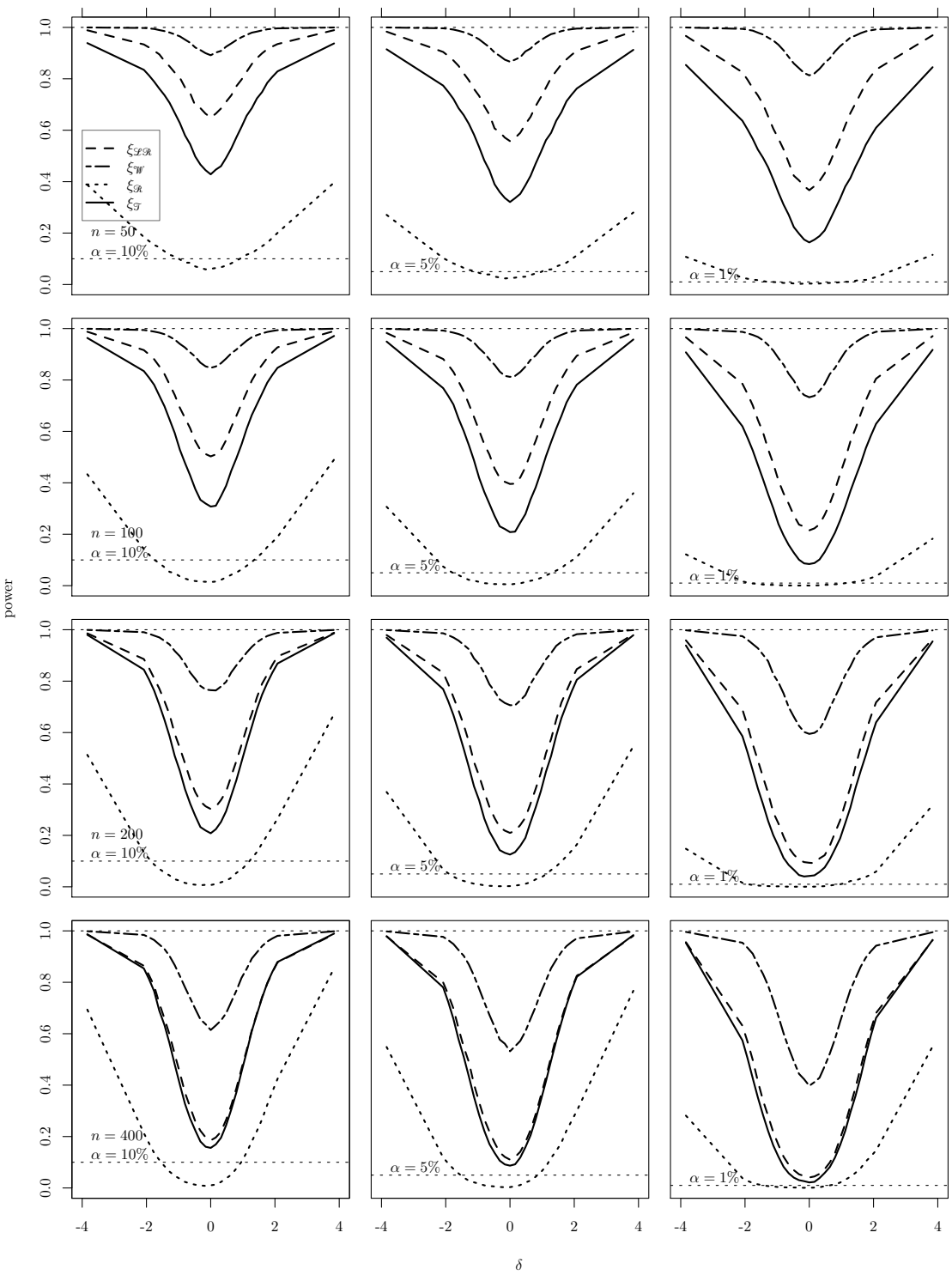


Figure 5.22: Non-null rejection rates of the four tests for gamma response model with NPML fitting and $K = 7$

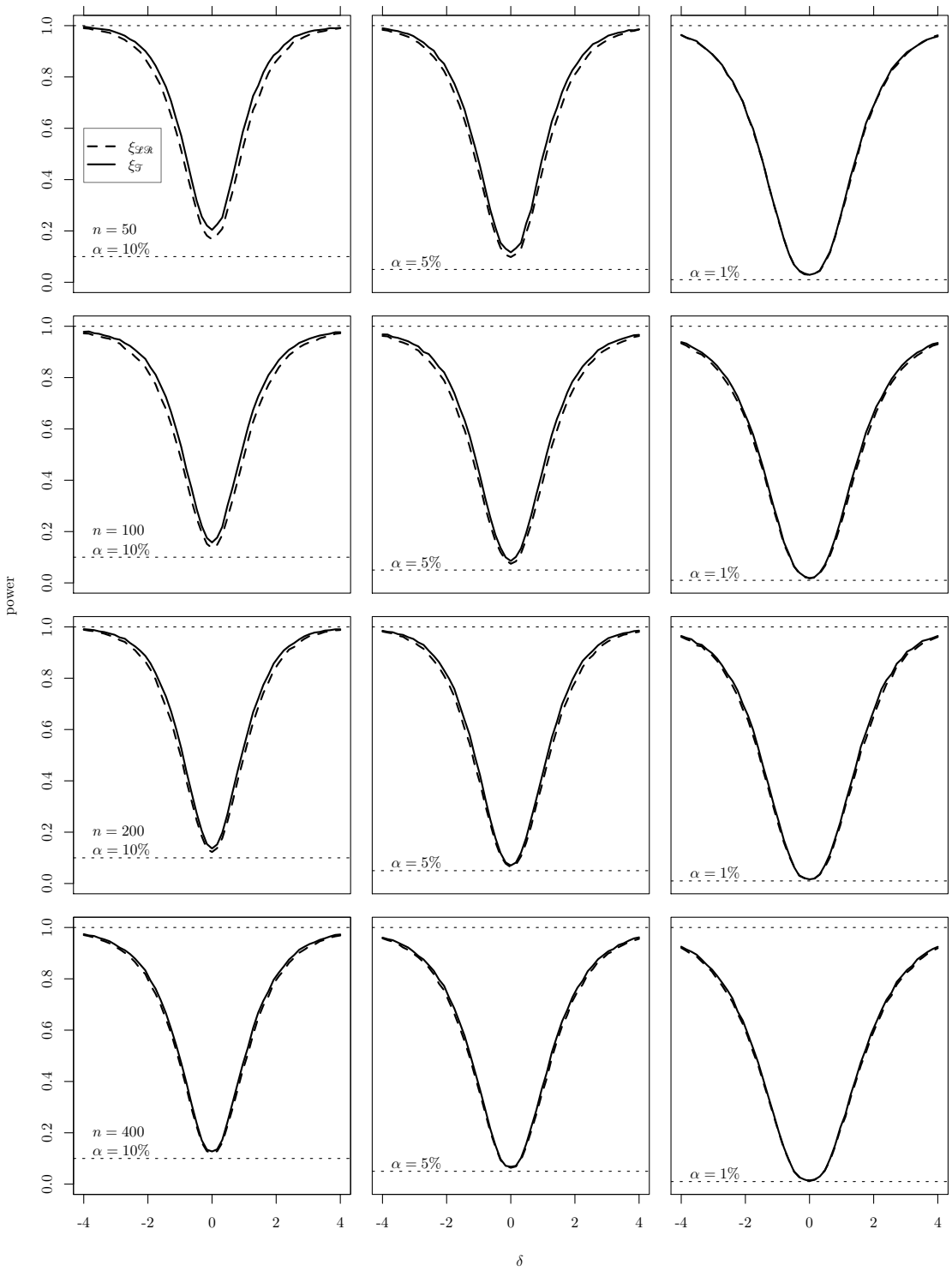


Figure 5.23: Non-null rejection rates of the four tests for gamma response variance components model with NPML fitting and $K = 3$

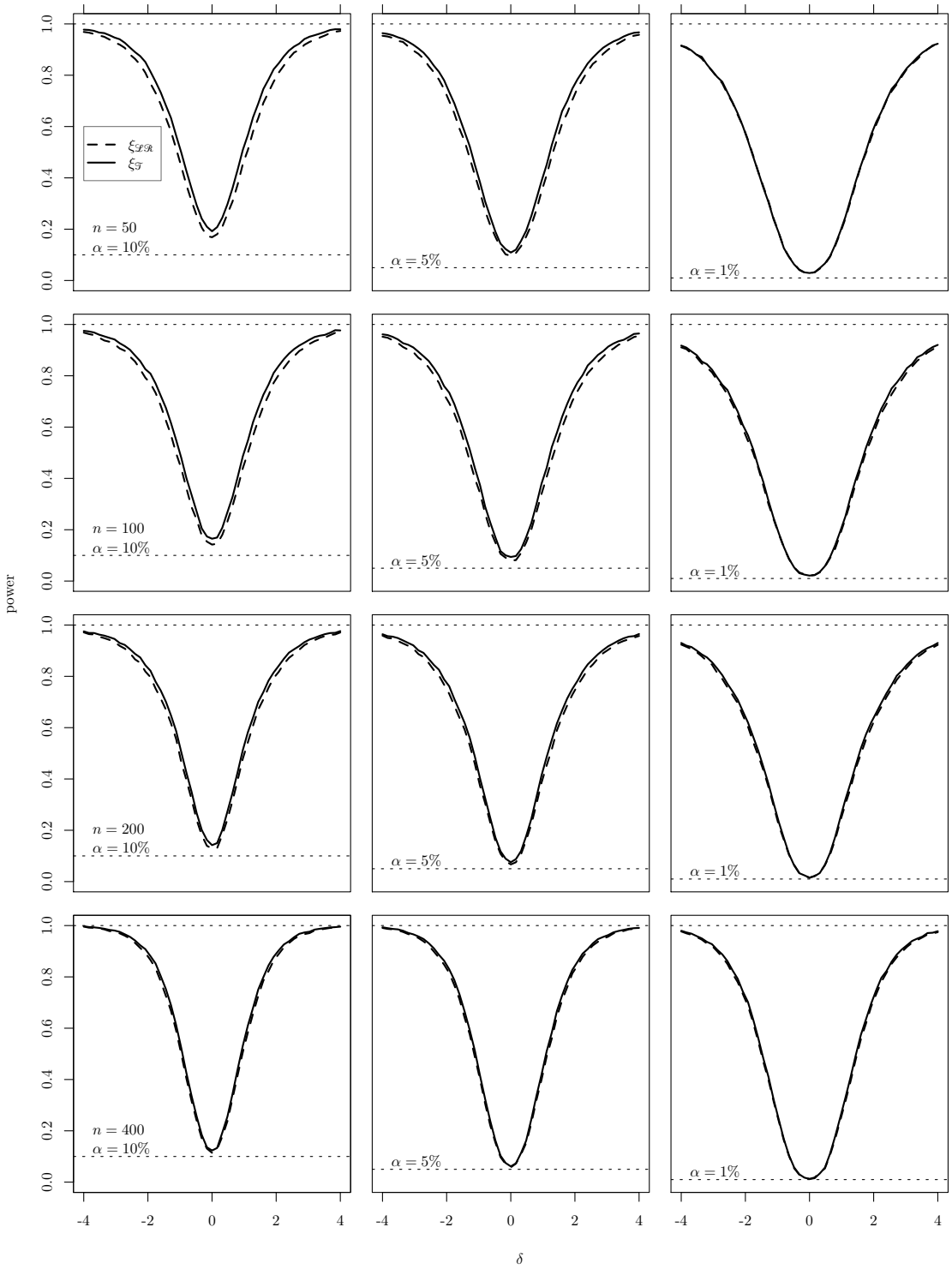


Figure 5.24: Non-null rejection rates of the four tests for gamma response variance components model with NPML fitting and $K = 5$

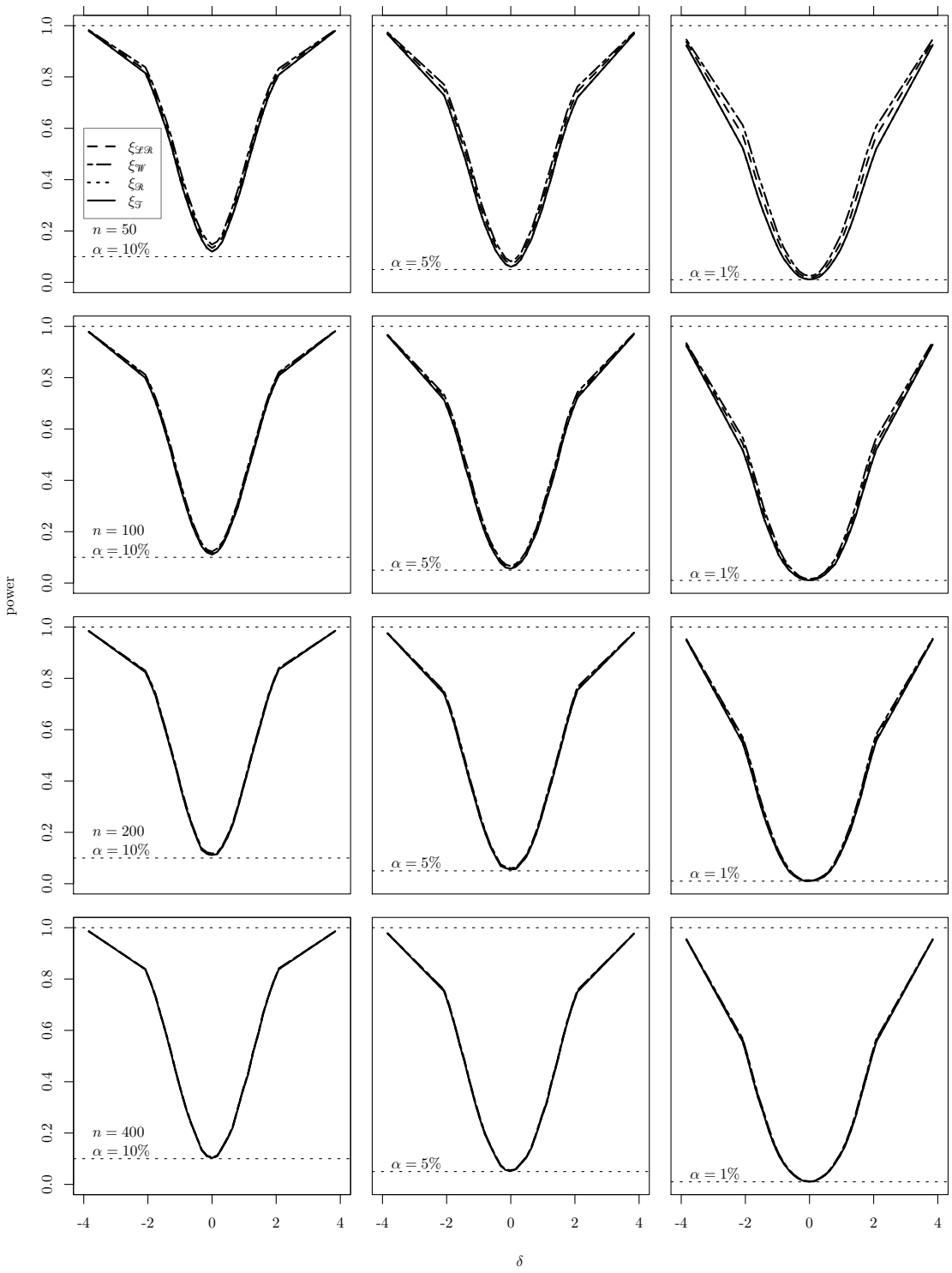


Figure 5.25: Non-null rejection rates of the four tests for normal response model with Gaussian quadrature fitting and $K = 3$

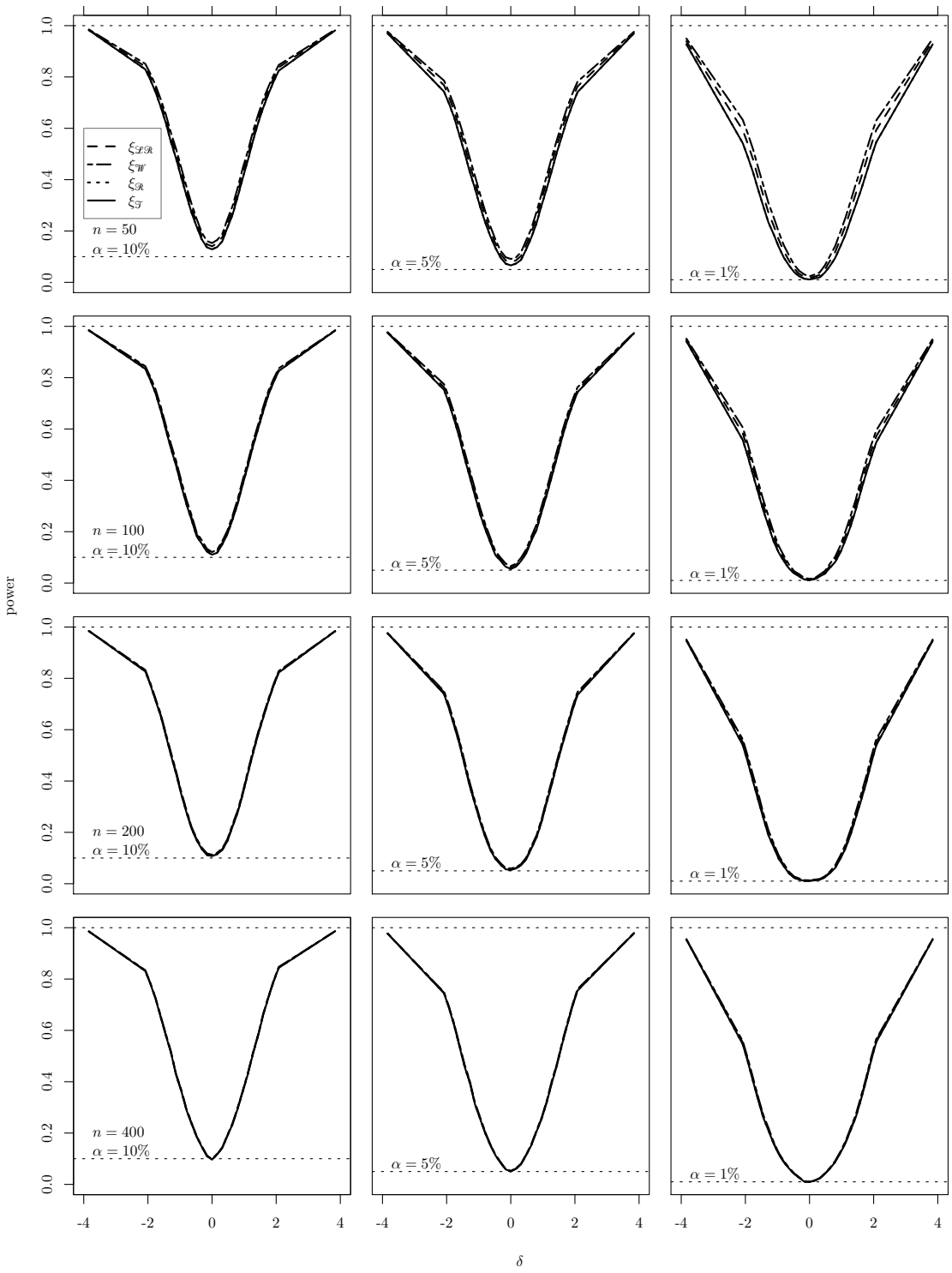


Figure 5.26: Non-null rejection rates of the four tests for normal response model with Gaussian quadrature fitting and $K = 5$

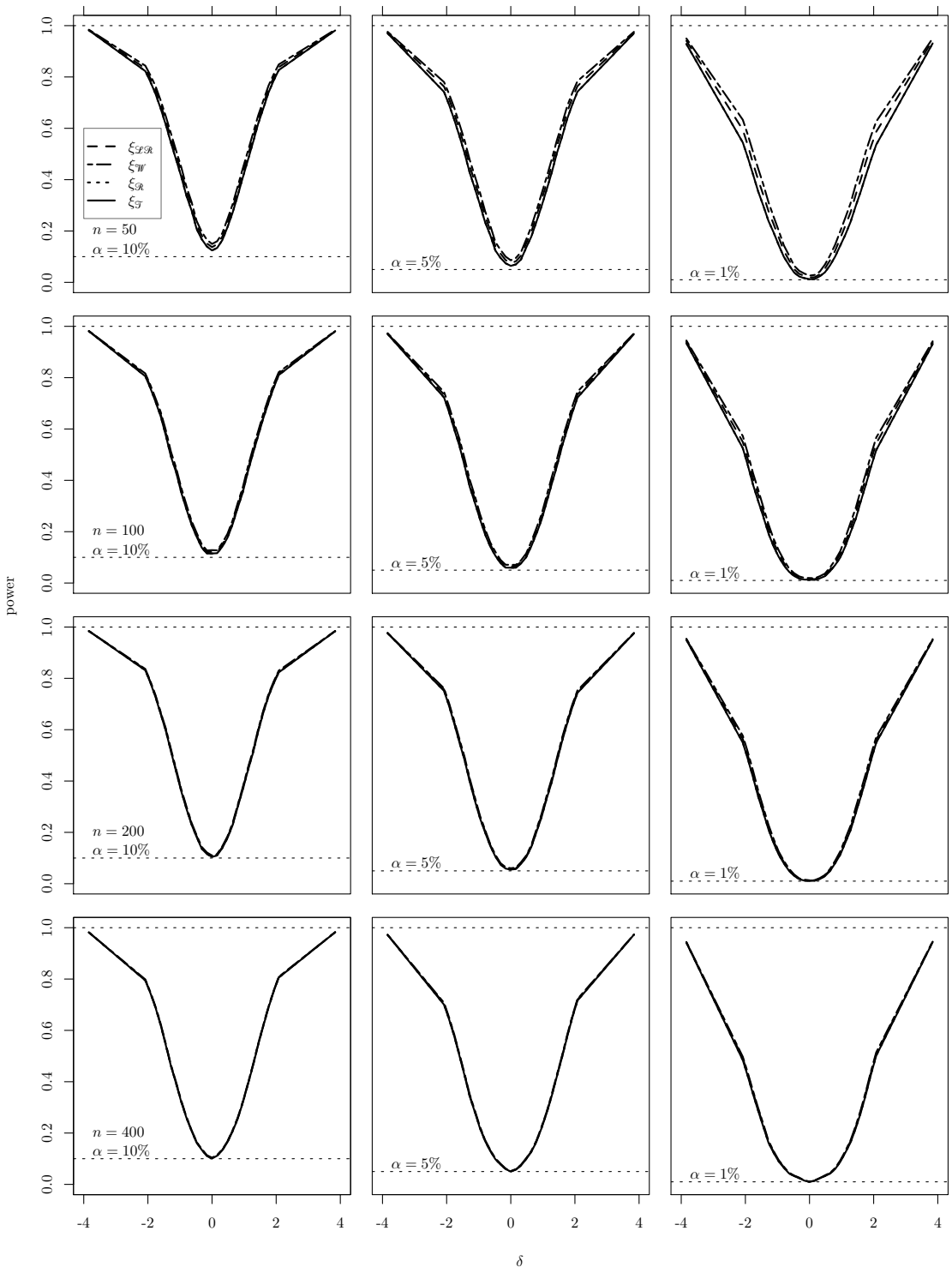


Figure 5.27: Non-null rejection rates of the four tests for normal response model with Gaussian quadrature fitting and $K = 7$

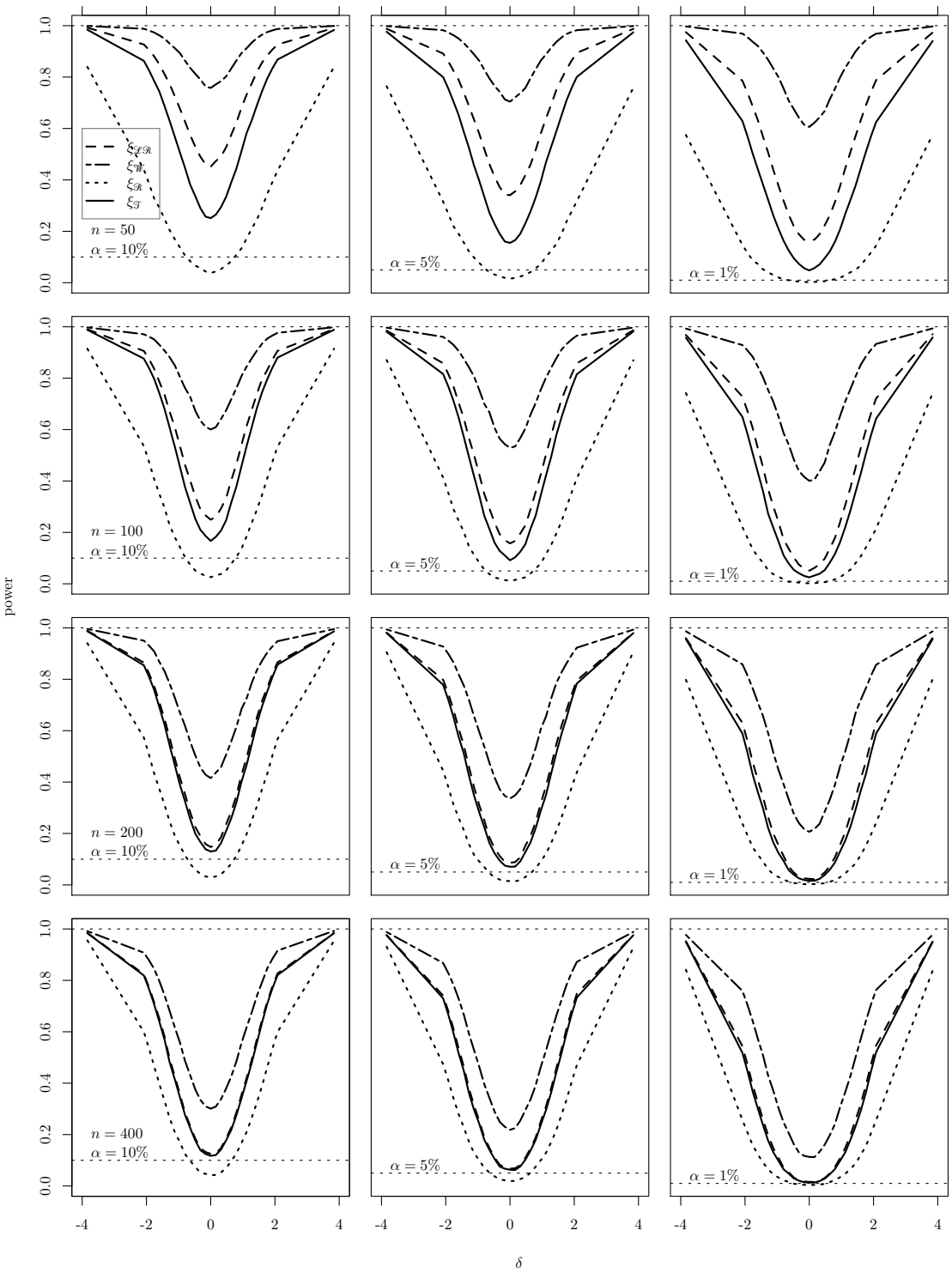


Figure 5.28: Non-null rejection rates of the four tests for normal response model with NPML fitting and $K = 3$

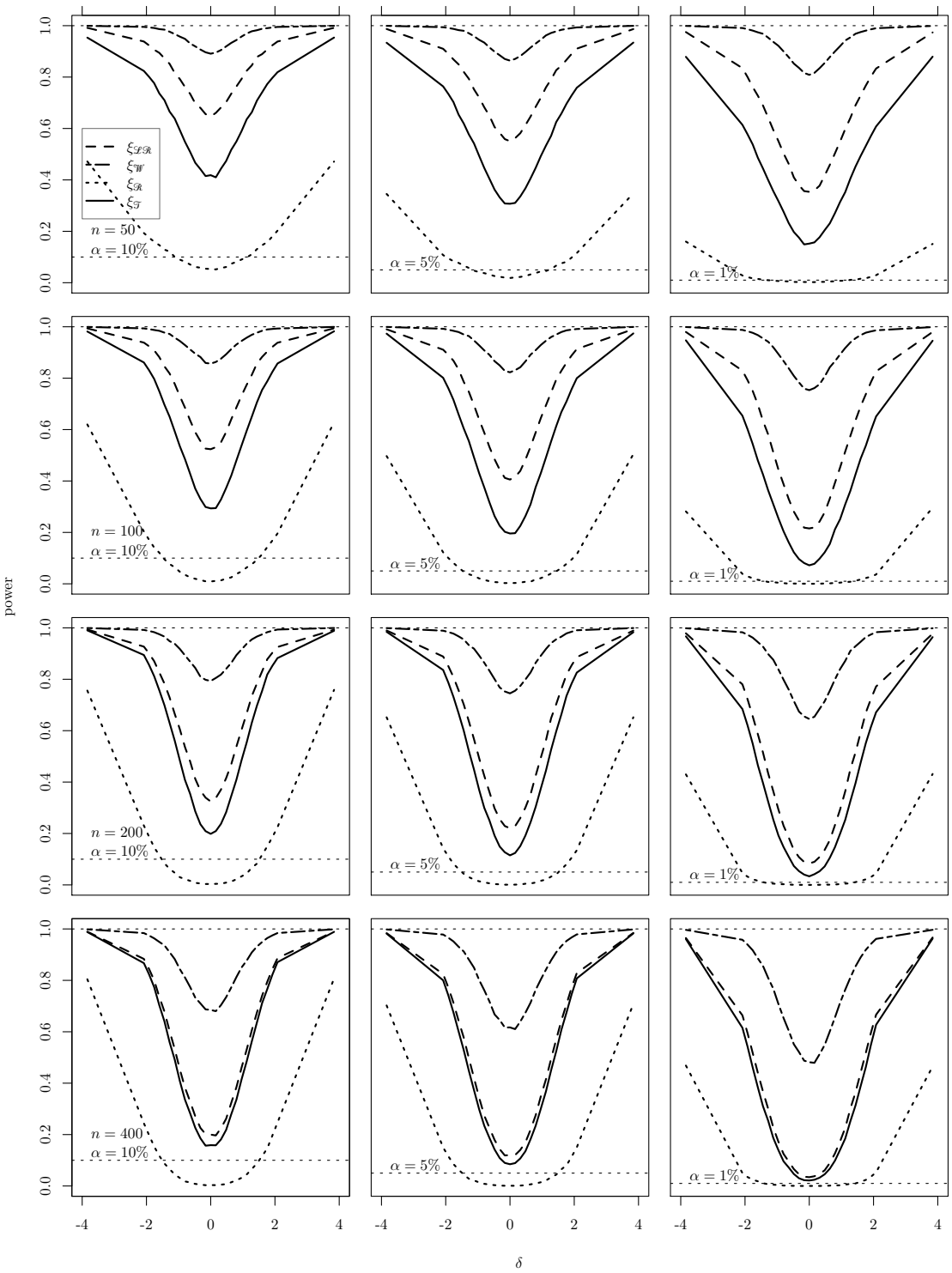


Figure 5.29: Non-null rejection rates of the four tests for normal response model with NPML fitting and $K = 5$

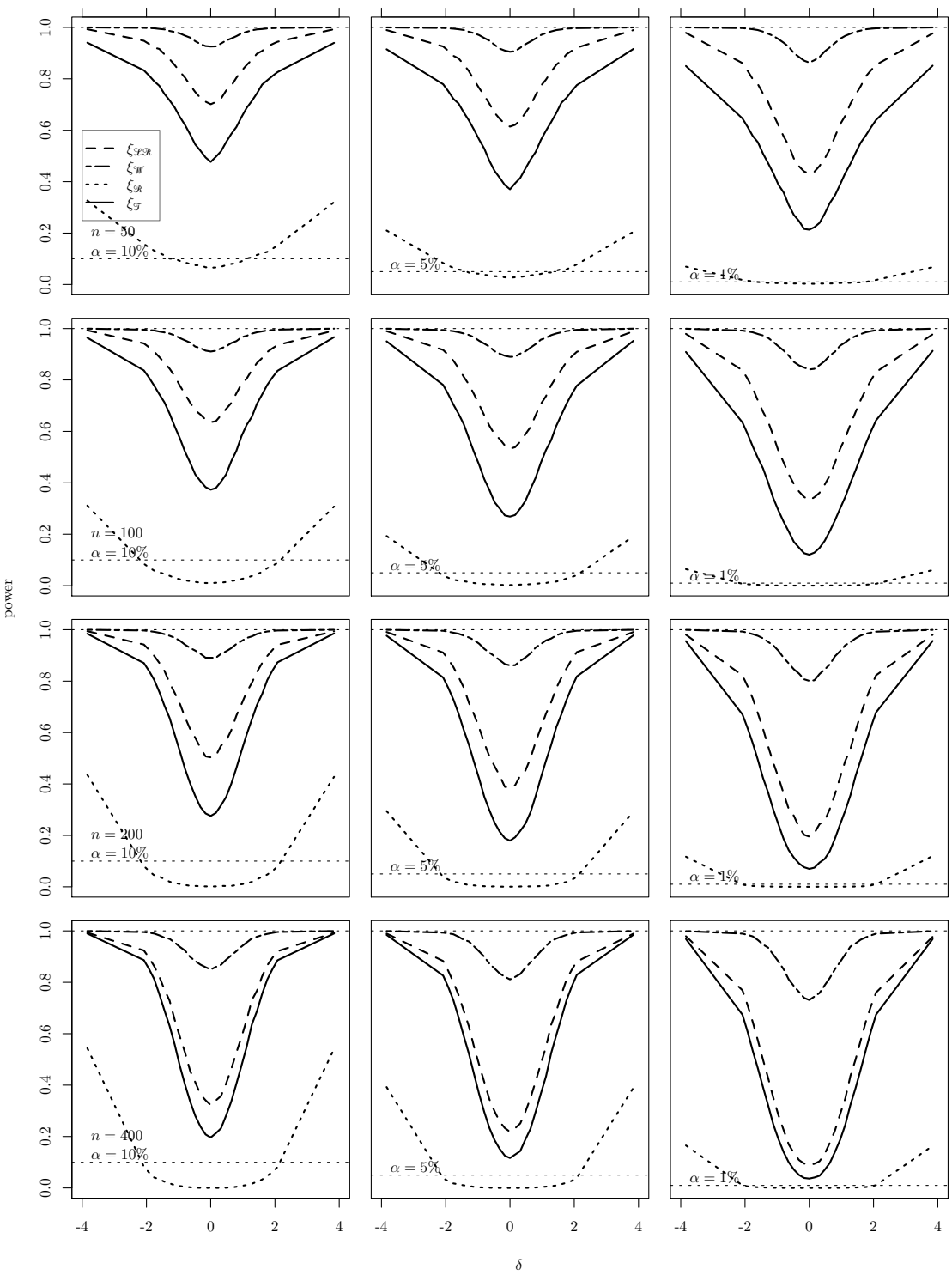


Figure 5.30: Non-null rejection rates of the four tests for normal response model with NPML fitting and $K = 7$

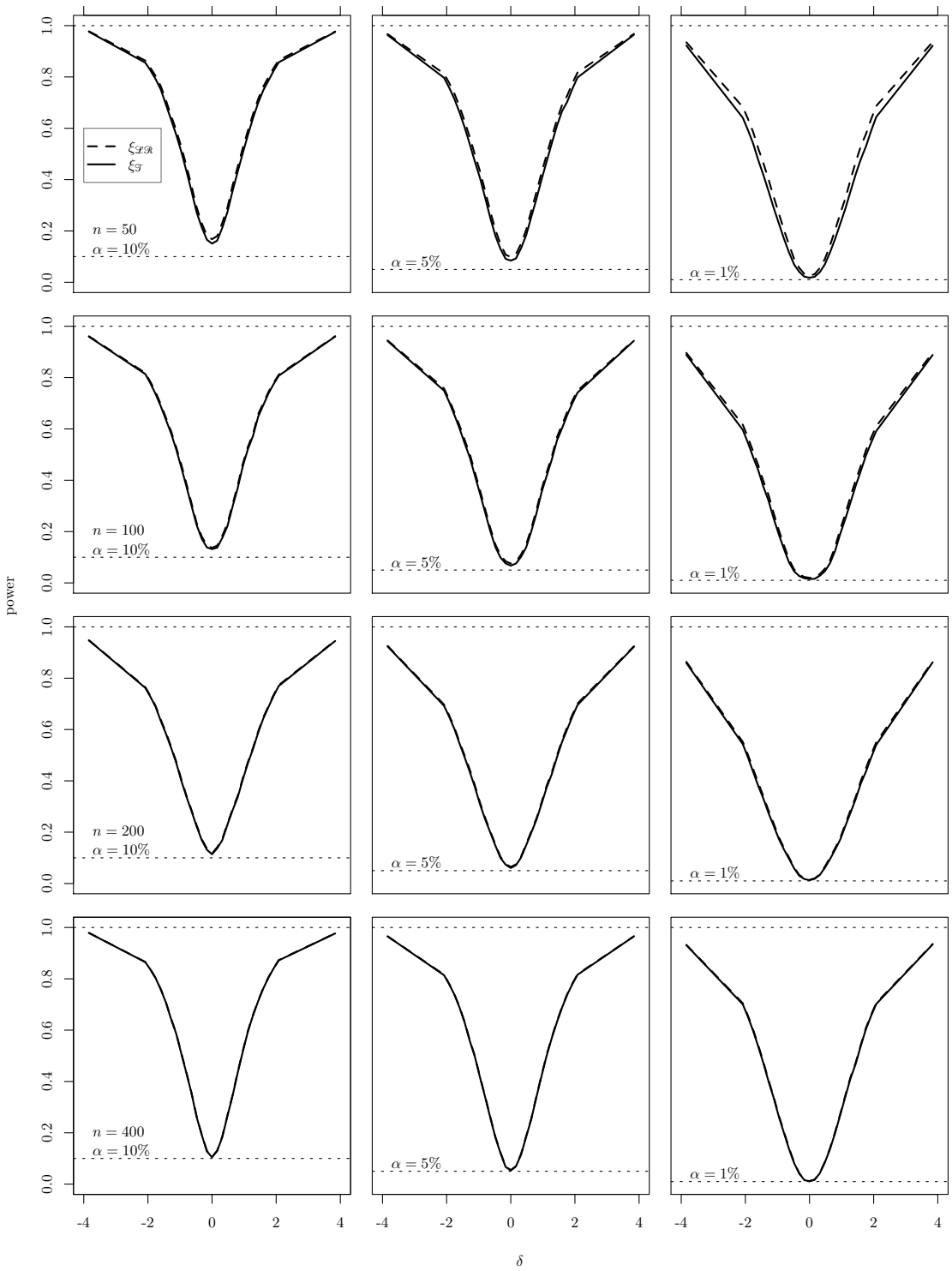


Figure 5.31: Non-null rejection rates of the four tests for normal response variance components model with NPML fitting and $K = 3$

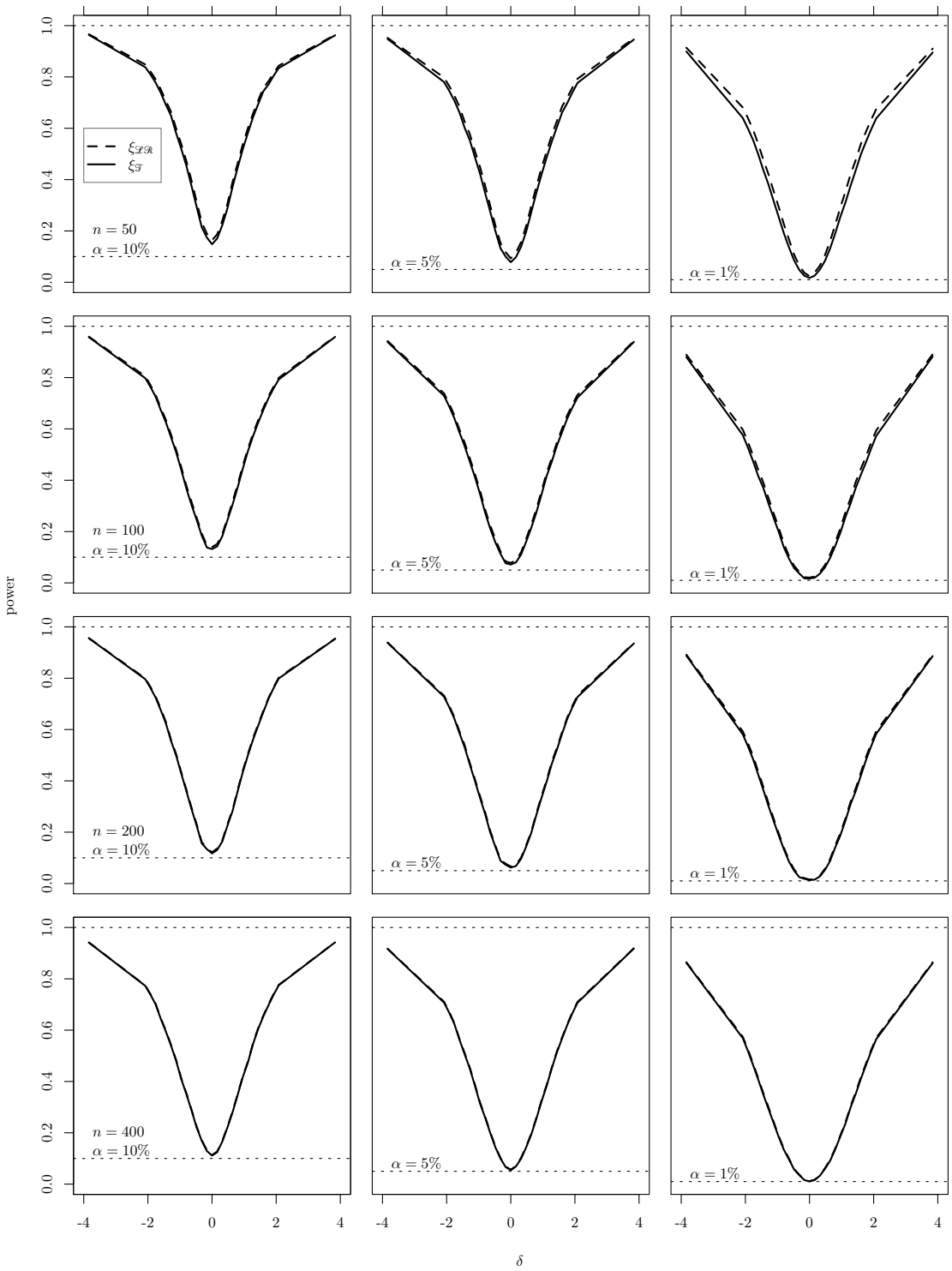


Figure 5.32: Non-null rejection rates of the four tests for normal response variance components model with NPML fitting and $K = 5$

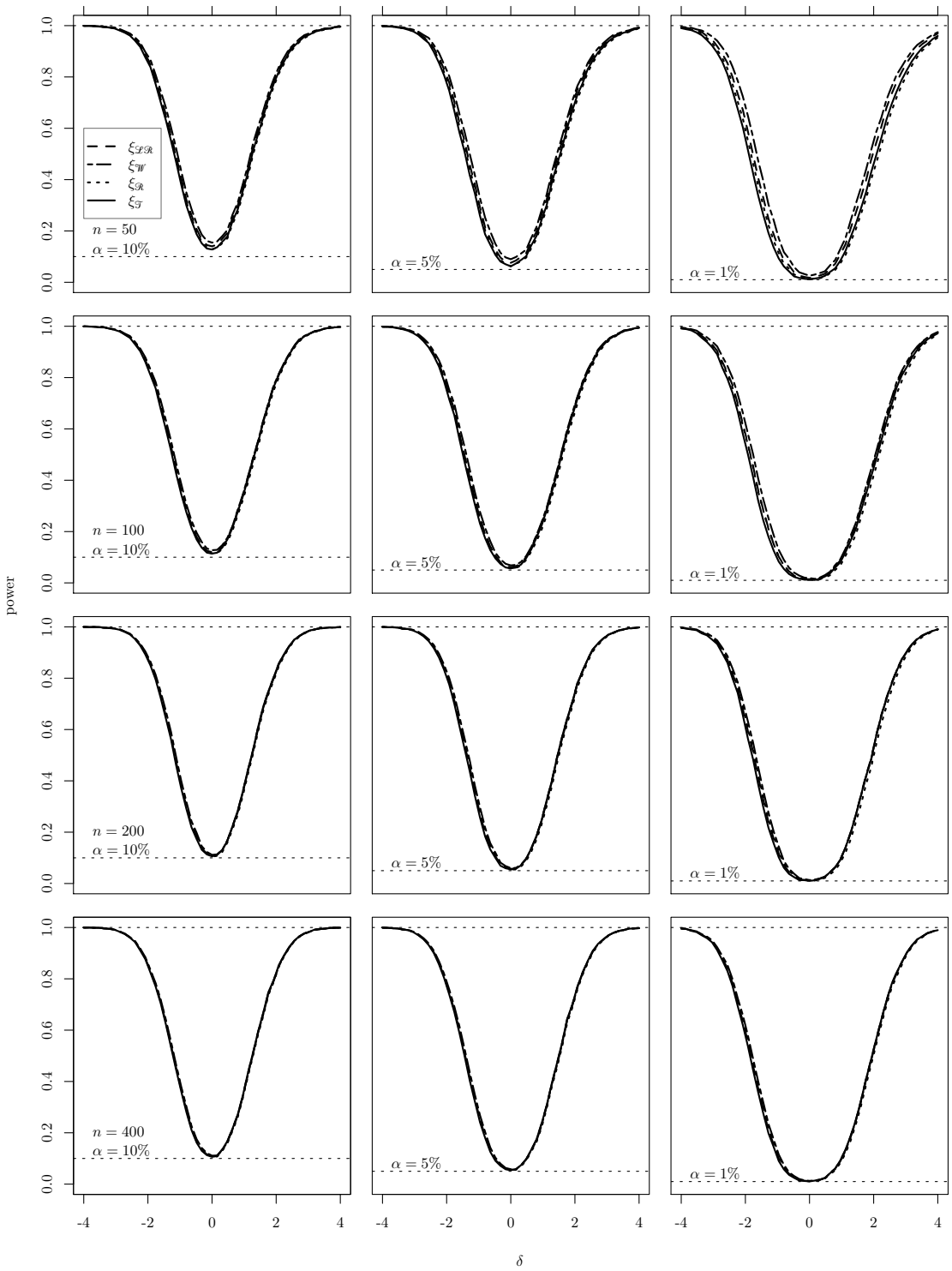


Figure 5.33: Non-null rejection rates of the four tests for inverse Gaussian response model with Gaussian quadrature fitting and $K = 3$

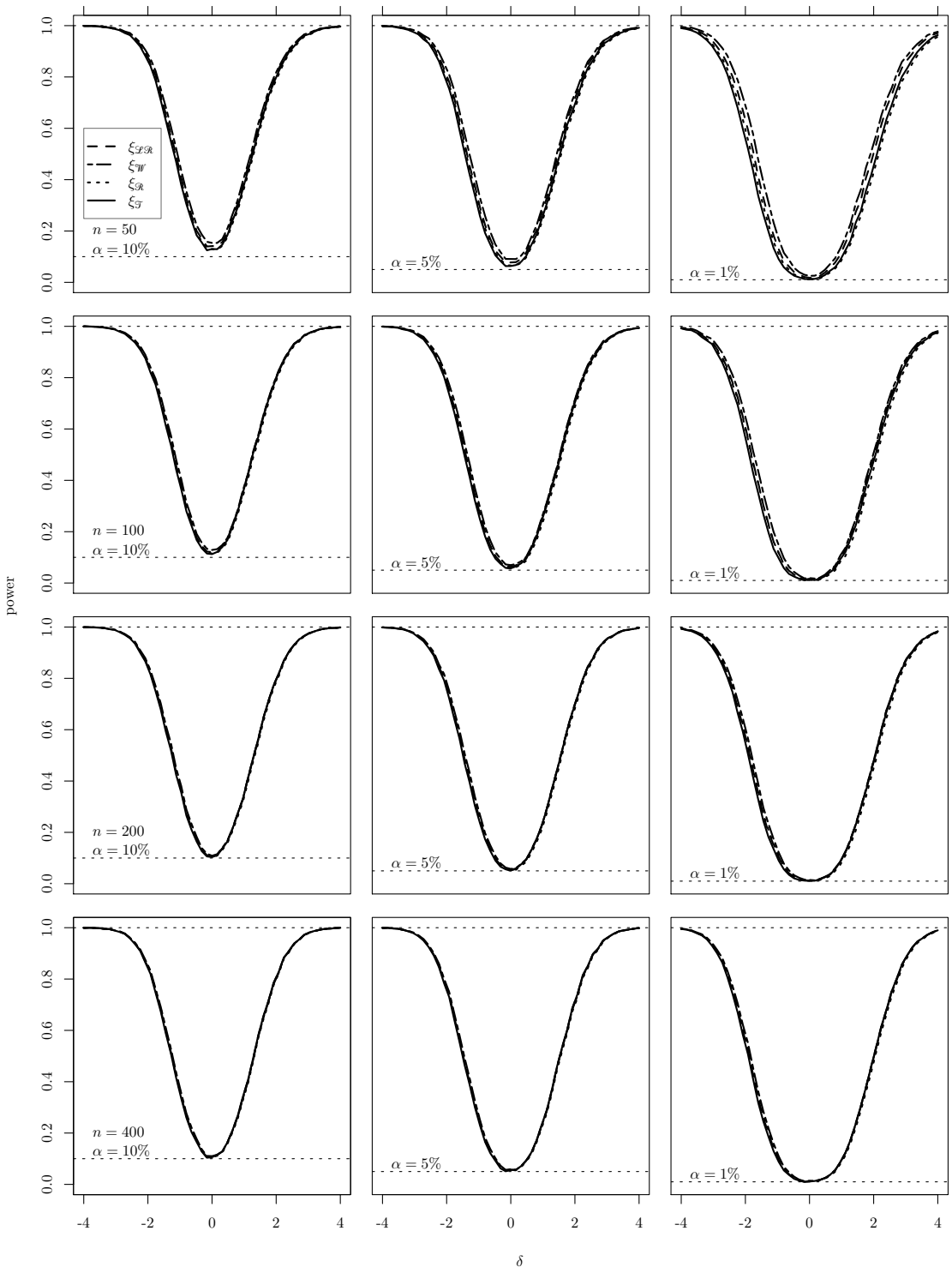


Figure 5.34: Non-null rejection rates of the four tests for inverse Gaussian response model with Gaussian quadrature fitting and $K = 5$

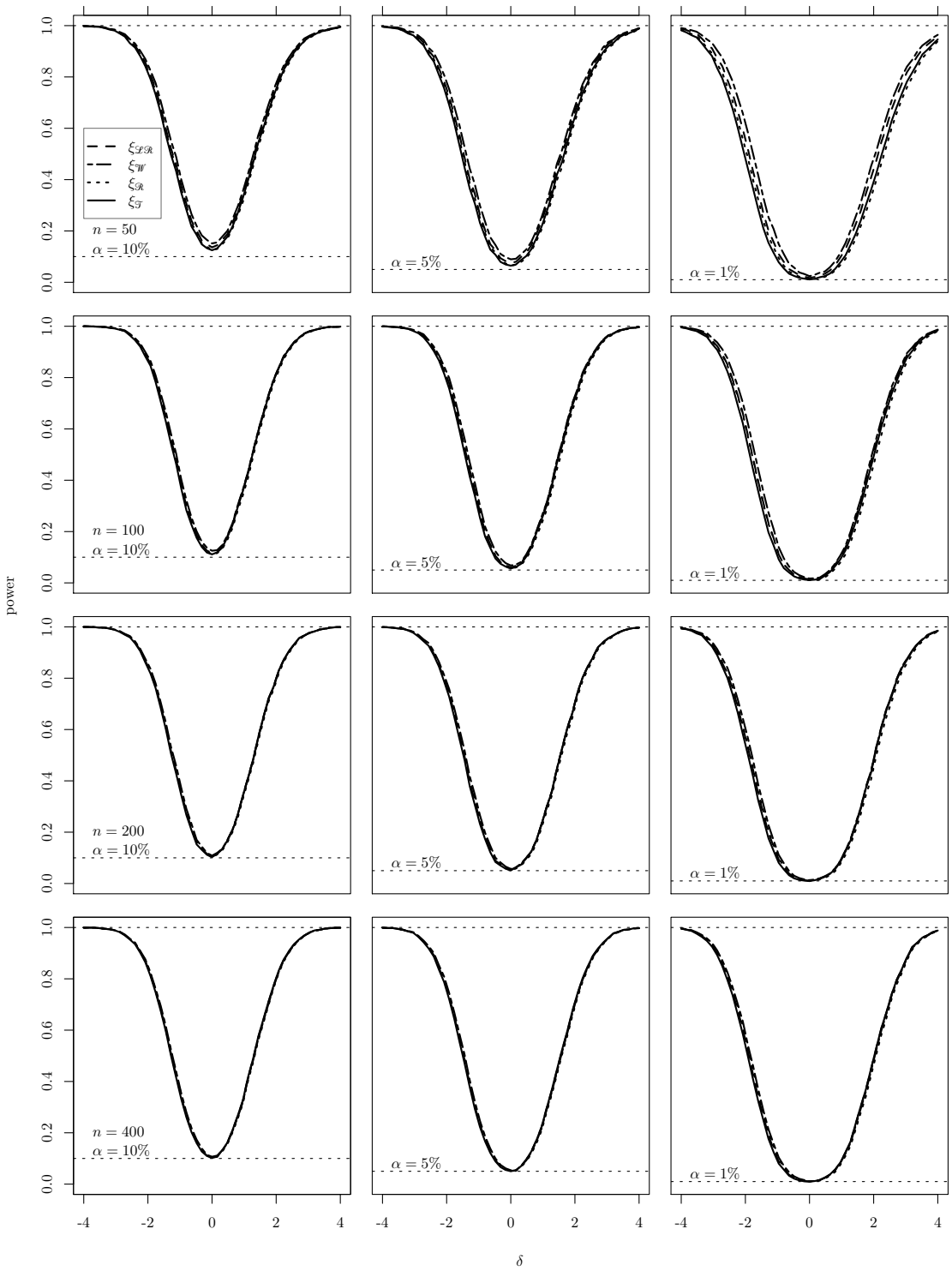


Figure 5.35: Non-null rejection rates of the four tests for inverse Gaussian response model with Gaussian quadrature fitting and $K = 7$

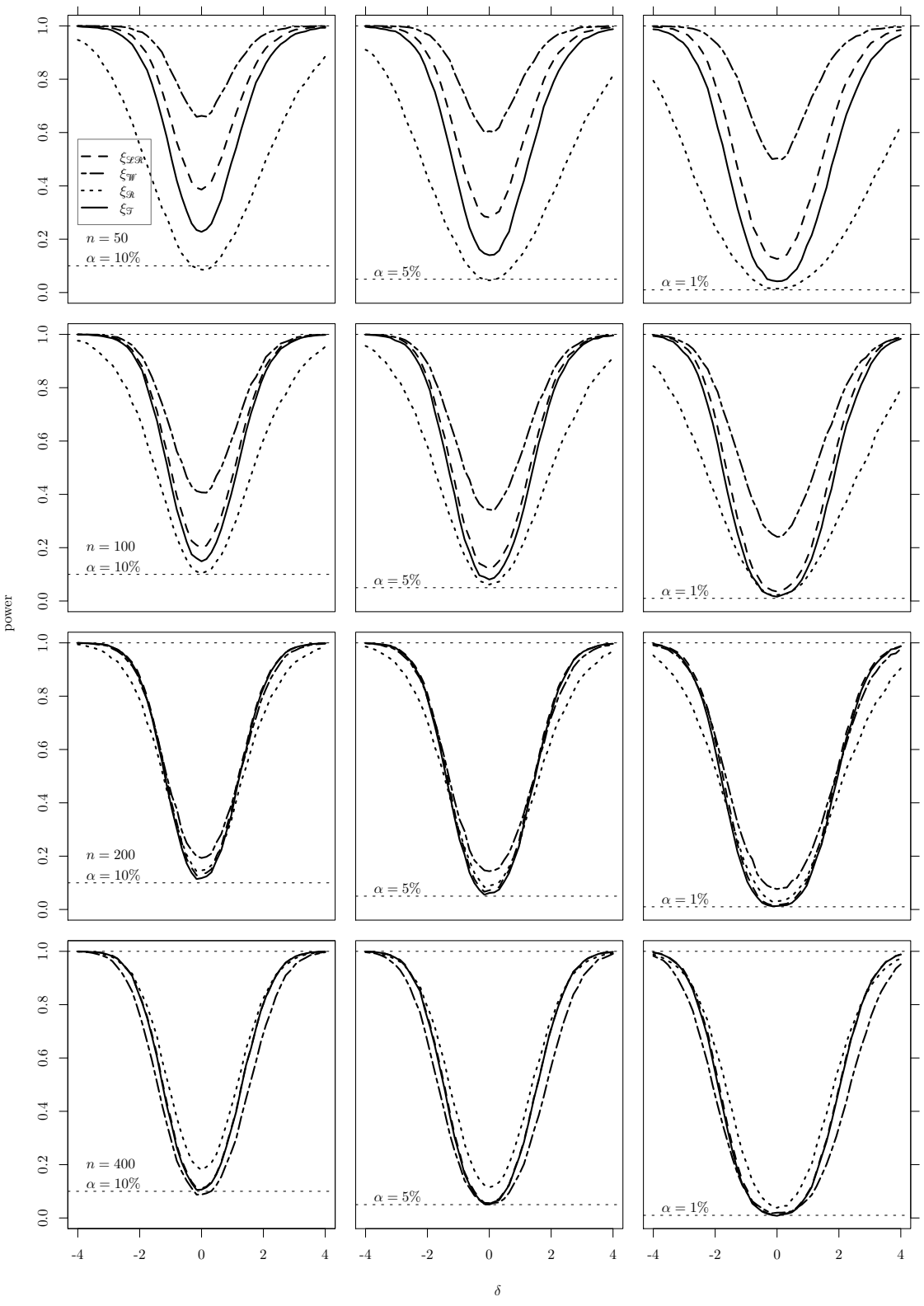


Figure 5.36: Non-null rejection rates of the four tests for inverse Gaussian response model with NPML fitting and $K = 3$

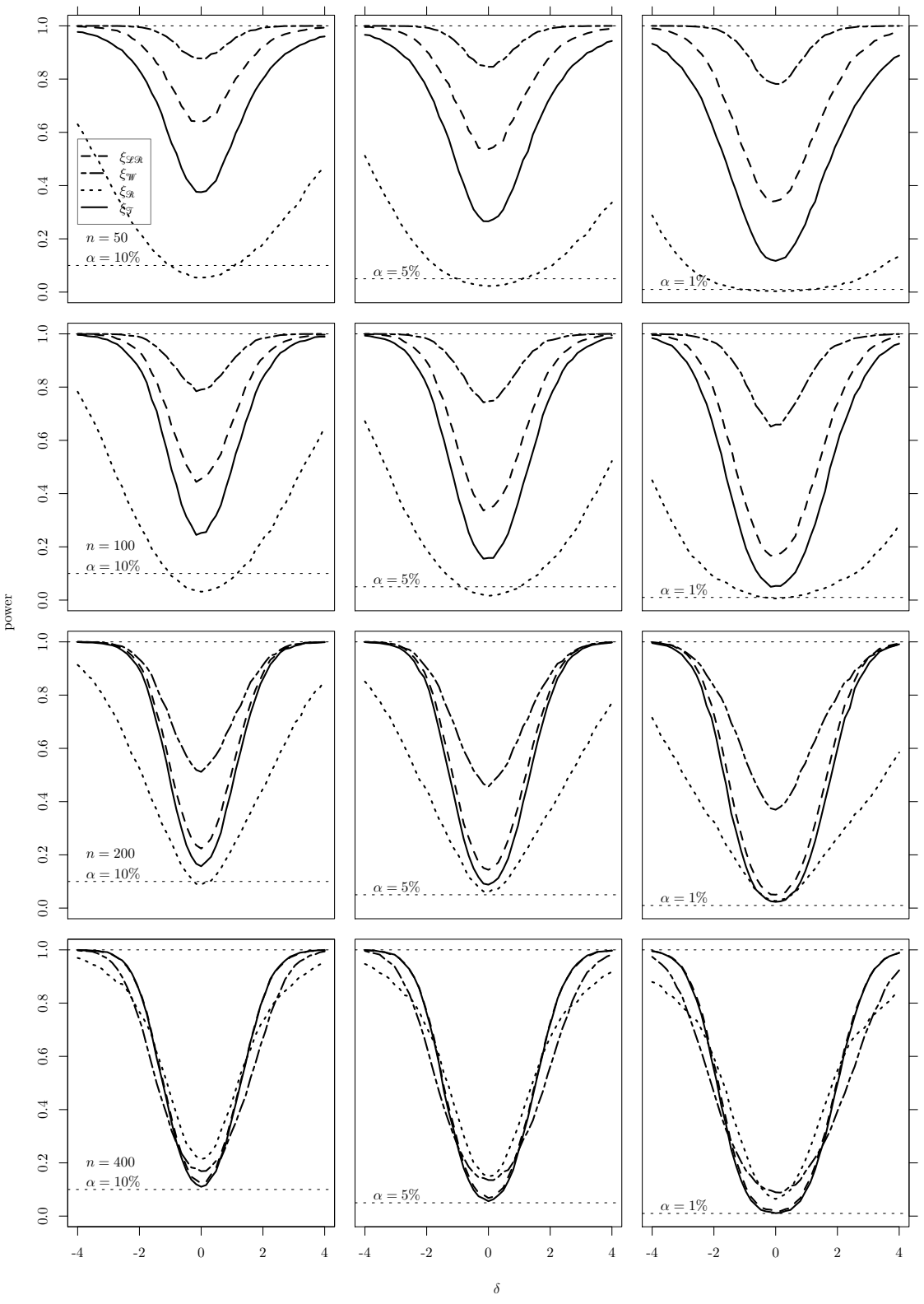


Figure 5.37: Non-null rejection rates of the four tests for inverse Gaussian response model with NPML fitting and $K = 5$

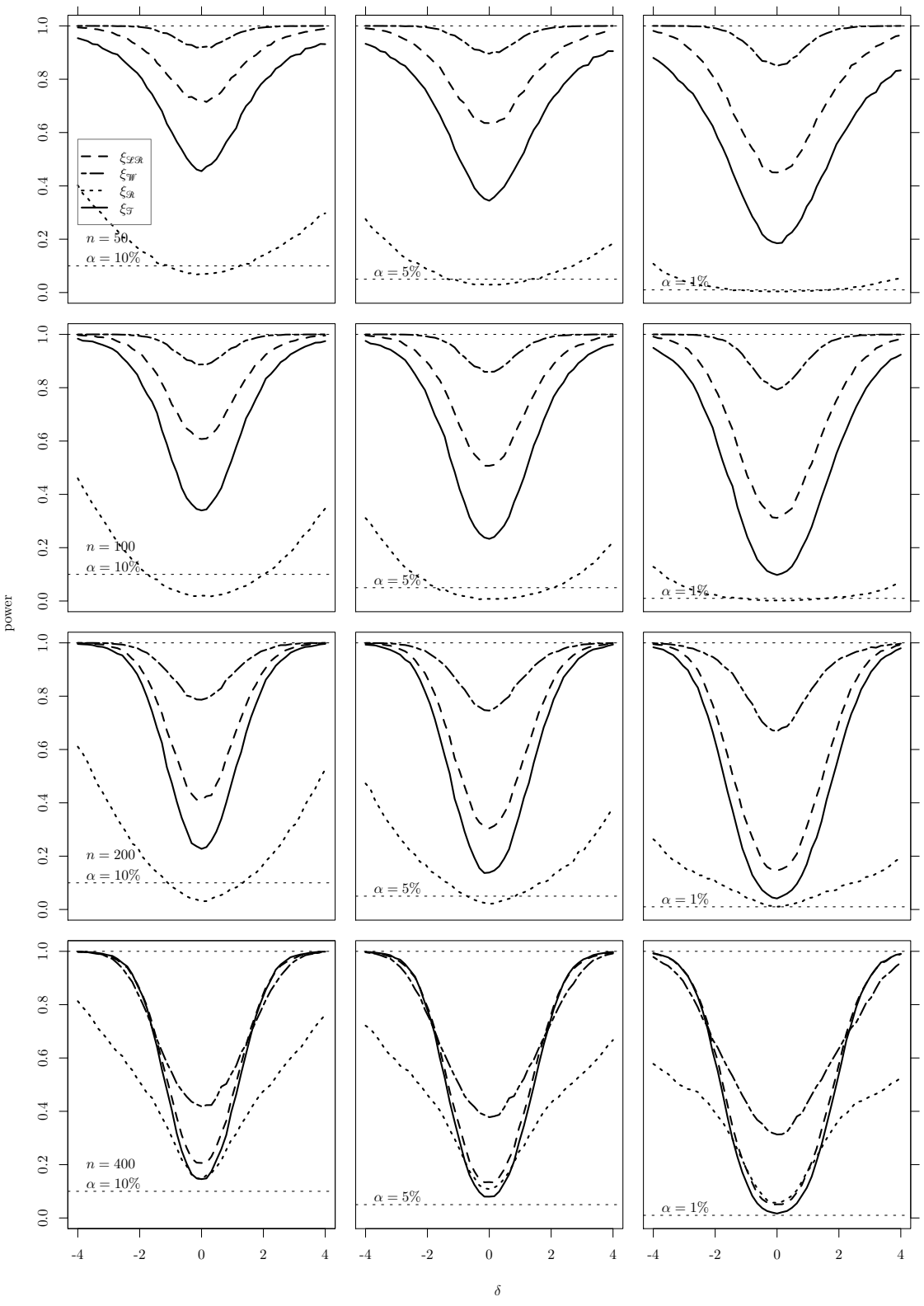


Figure 5.38: Non-null rejection rates of the four tests for inverse Gaussian response model with NPML fitting and $K = 7$

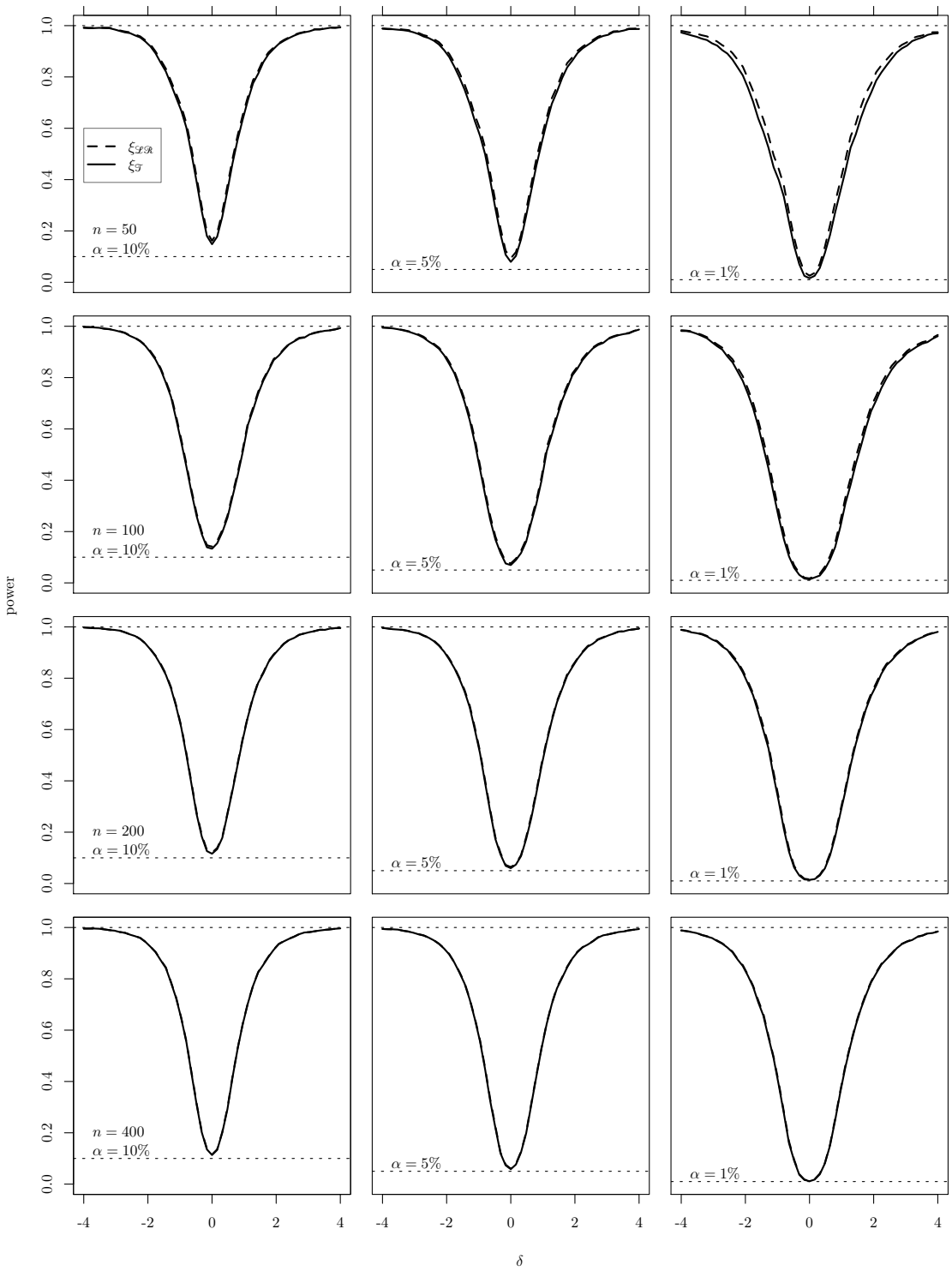


Figure 5.39: Non-null rejection rates of the four tests for inverse Gaussian response variance components model with NPML fitting and $K = 3$

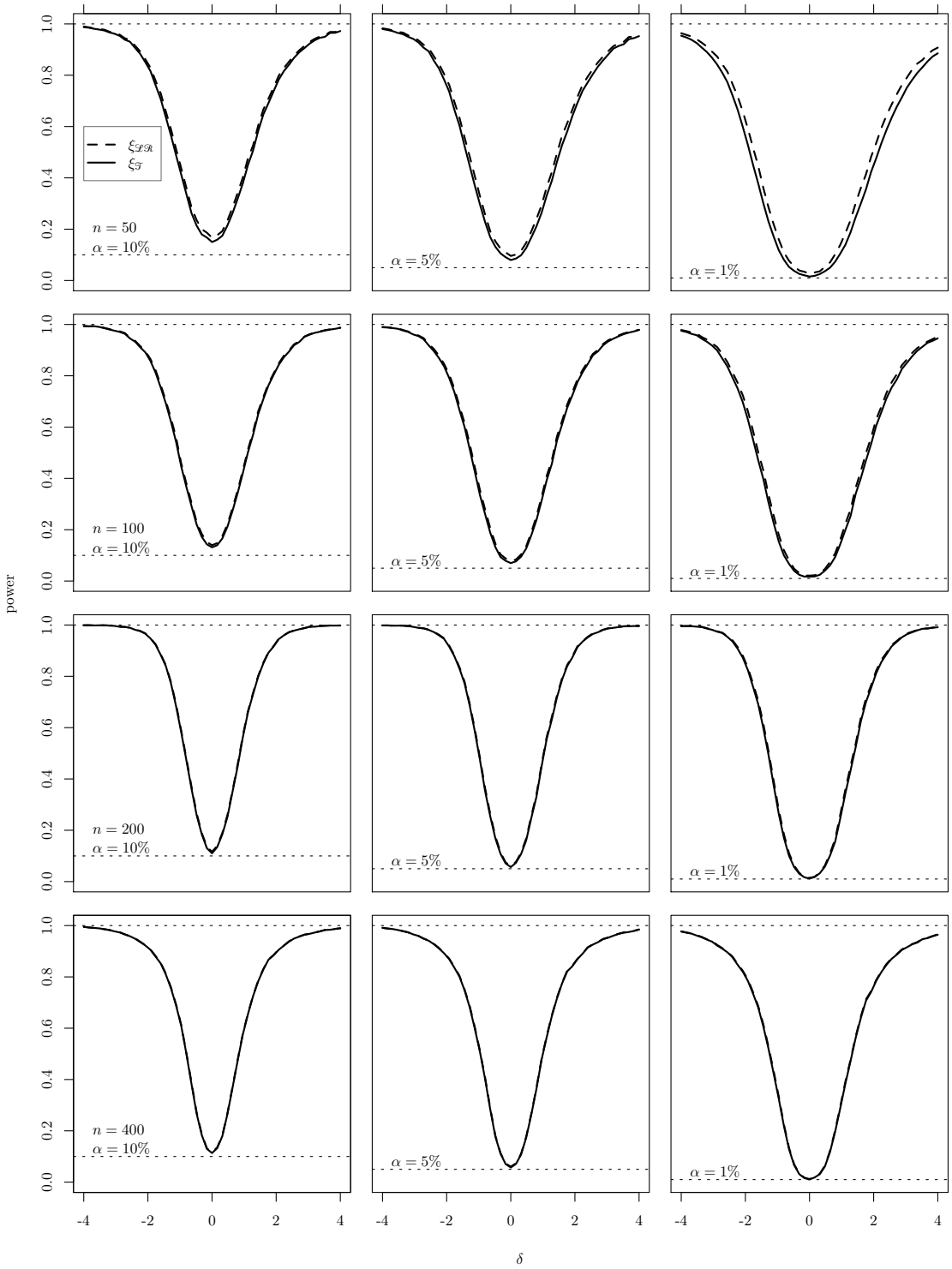


Figure 5.40: Non-null rejection rates of the four tests for inverse Gaussian response variance components model with NPML fitting and $K = 5$

## 5.2 Real data examples

We now provide four examples to illustrate the application of the gradient test. All examples were performed using the code provided in Appendix A. Any code provided in later subsections must be preceded by the following lines in R.

```
require(npmlreg)

## Loading required package:  npmlreg

require(Matrix)

## Loading required package:  Matrix

source("lr.test.R")
source("wald.test.R")
source("rao.test.R")
source("gradient.test.R")
```

The first two lines in this code load the packages **npmlreg** (Einbeck et al., 2014) and **Matrix** (Bates & Maechler, 2017), respectively. The first package is needed for fitting any GLMwRE presented in this section and the latter is required because `wald.test`, `rao.test` and `gradient.test` have embedded functions from this package. The last four lines load self-written functions to compute the likelihood ratio test (`lr.test`), Wald test (`wald.test`), Rao test (`rao.test`) and gradient test (`gradient.test`). All the test functions must be stored in the respective `*.R` files.
All four functions require the same basic arguments `fit.null` and `subset.formula`. The argument `fit.null` receives the object resulting from the model fitted under the null hypothesis using either `alldist` or `allvc`. The argument `subset.formula` receives the formula corresponding to the subset of covariates under test. For instance, supose a linear predictor for the full model such as

$$\eta_i = \beta_0 + \beta_1 \texttt{x1}_i + \beta_2 \texttt{x2}_i + \beta_3 \texttt{x3}_i + \beta_4 \texttt{x4}_i + z_i^*,$$

and we would like to test $\mathscr{H}_0 : \beta_3 = \beta_4 = 0$, then we should fit the null model

```
# for simple overdispersion model...
fit.null <- alldist(y  ~ x1 + x2,
                    k = k,
                    data = data,
                    family = family,
                    random.distribution = "gq") # or "np"
# ... or for variance components model
fit.null <- allvc(y  ~ x1 + x2,
                  random = ~ 1|id,
                  k = k,
                  data = data,
                  family = family,
                  random.distribution = "gq") # or "np"
```

where the user must inform the appropriate arguments for `random`, `k`, `data`, `family` and `random.distribution`. After that, we are able to use the test functions with the code shown below.

```
lr.test(fit.null, ~ x3 + x4)
wald.test(fit.null, ~ x3 + x4)
rao.test(fit.null, ~ x3 + x4)
gradient.test(fit.null, ~ x3 + x4)
```

Each test function returns the values `statistic`, `parameter` and `p.value` which correspond to the test statistic, degrees of freedom and the $p$ value. This is also show in R like this.

```
##
##  Gradient test for GLMwRE
##
##  null model: y ~ x1 + x2
##  alt. model: y ~ x1 + x2 + x3 + x4
```

```
##
## statistic = 4.4963, parameter = 2, p-value = 0.1056
```

We point out that if the argument `random.distribution = "gq"` in `alldist` then the default variance estimation is performed using the analytic formulae presented in Section 4.2. This is equivalent to using `analytic.var = TRUE` however one can choose EM variance estimation simply by changing for `analytic.var = FALSE`. For `random.distribution = "np"`, the argument `analytic.var` is irrelevant because in theory we are only able to use the EM variance estimate. Because `wald.test` and `rao.test` rely on the Fisher information or information matrices and we restrict ourselves to developing the corresponding theory only for the classic overdispersion models, these two functions do not have an explicit implementation for variance components models and therefore we advise to not make use of them for `fit.null` fitted by `allvc`.

### 5.2.1 Risk factors for endometrial cancer grade

This dataset concerns the histology grade and risk factors for 79 cases of endometrial cancer which can be found in Heinze & Schemper (2002); it has a detailed description in Agresti (2015, Section 5.7) and is fully available in the package **brglm2** (Kosmidis, 2017) through `data(endometrial)`. This data includes the variables

`HG` histology of 79 cases (0 = low grade for 30 patients, 1 = high grade for 49 patients)

`NV` neovasculation (1 = present for 13 patients, 0 = absent for 66 patients)

`PI` pulsatility index of arteria uterina (ranging from 0 to 49)

`EH` endometrium height (ranging from 0.27 to 3.61).

The original analysis presented by Heinze & Schemper (2002) uses a logistic regression model for $\mu_i = \mathrm{E}[\mathtt{HG}]$ with linear predictor

$$\mathrm{logit}(\mu_i) = \beta_0 + \beta_1 \mathtt{NV}_i + \beta_2 \mathtt{NI}_i + \beta_3 \mathtt{EH}_i, \quad \text{for} \quad i = 1, \ldots, 79. \tag{5.2.9}$$

where $\mu_i = \mathrm{E}[\texttt{HG}]$, $\text{logit}(\mu_i) = \log(\mu_i) - \log(1-\mu_i)$ and $\beta_0$, $\beta_1$, $\beta_2$ and $\beta_3$ are unknown parameters.

For our analysis, we include a random effect $z_i$ to (5.2.9) such as

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \texttt{NV}_i + \beta_2 \texttt{NI}_i + \beta_3 \texttt{EH}_i + \sigma z_i, \quad \text{for} \quad i = 1, \ldots, 79. \tag{5.2.10}$$

where $\sigma > 0$ is unknown. We supose $z_i \sim \mathcal{N}(0,1)$ for estimation purposes and we choose $k = 4$. In R, we can load the data and fit the model (5.2.10) using the following code.

```
require(brglm2)

## Loading required package: brglm2

data(endometrial) # load data

fit.null = alldist(HG ~ NV + PI + EH, data = endometrial,
                   family = binomial(logit), k = 4,
                   random.distribution = "gq",
                   plot.opt = 0, verbose = FALSE)

summary(fit.null)

##
## Call:  alldist(formula = HG ~ NV + PI + EH,
##                family = binomial(logit), data = endometrial,
##                k = 4, random.distribution = "gq",
##                plot.opt = 0, verbose = FALSE)
##
## Coefficients:
##                Estimate   Std. Error    t value
## (Intercept)  4.30816829 1.638445e+00  2.6294243
## NV          18.18796847 1.714746e+03  0.0106068
```

```
## PI          -0.04218358 4.434296e-02 -0.9513026
## EH          -2.90566755 8.463923e-01 -3.4330034
## z            0.08701383 3.346314e-01  0.2600289
##
## Random effect distribution - standard deviation:    0.08701383
##
##
## -2 log L:     55.4      Convergence at iteration  23
```

Supose one would like to test if the quadratic value of `PI` has a relevant effect in the model. The alternative linear predictor is then expressed as

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \texttt{NV}_i + \beta_2 \texttt{NI}_i + \beta_3 \texttt{EH}_i + \beta_4 \texttt{PI}_i^2 + \sigma z_i, \quad \text{for} \quad i = 1, \ldots, 79.$$

This is equivalent to test the hypothesis

$$\begin{cases} \mathscr{H}_0 : \beta_4 = 0 \\ \mathscr{H}_1 : \beta_4 \neq 0 \end{cases}$$

which can be numerically evaluated by the following R code.

```
lr.test(fit.null,~I(PI^2))

##
##  Likelihood ratio test for GLMwRE
##
##  null model: HG ~ NV + PI + EH
##  alt. model: HG ~ NV + PI + EH + I(PI^2)
##
## statistic = 9.0784, parameter = 1, p-value = 0.002586

wald.test(fit.null,~I(PI^2))

##
```

```
##  Wald test for GLMwRE (by analytic variance)
##
##  null model: HG ~ NV + PI + EH
##  alt. model: HG ~ NV + PI + EH + I(PI^2)
##
## statistic = 7.3862, parameter = 1, p-value = 0.006573

wald.test(fit.null,~I(PI^2),analytic.var=FALSE)

##
##  Wald test for GLMwRE (by EM variance estimate)
##
##  null model: HG ~ NV + PI + EH
##  alt. model: HG ~ NV + PI + EH + I(PI^2)
##
## statistic = 7.1336, parameter = 1, p-value = 0.007565

rao.test(fit.null,~I(PI^2))

##
##  Rao test for GLMwRE (by analytic variance)
##
##  null model: HG ~ NV + PI + EH
##  alt. model: HG ~ NV + PI + EH + I(PI^2)
##
## statistic = 14.405, parameter = 1, p-value = 0.0001474

rao.test(fit.null,~I(PI^2),analytic.var=FALSE)

##
##  Rao test for GLMwRE (by EM variance estimate)
##
##  null model: HG ~ NV + PI + EH
##  alt. model: HG ~ NV + PI + EH + I(PI^2)
```

```
##
## statistic = 14.403, parameter = 1, p-value = 0.0001476

gradient.test(fit.null,~I(PI^2))

##
##  Gradient test for GLMwRE
##
##  null model: HG ~ NV + PI + EH
##  alt. model: HG ~ NV + PI + EH + I(PI^2)
##
## statistic = 9.1896, parameter = 1, p-value = 0.002434
```

The Table 5.12 summarises the results for the four tests.

Table 5.12: Results for testing $\mathscr{H}_0 : \beta_4 = 0$

| Statistic | value | $p$ value |
|---|---|---|
| $\xi_{\mathscr{LR}}$ | 9.0784 | 0.002586 |
| $\xi_{\mathscr{W}}$ | 7.3862 | 0.006573 |
| $\xi^*_{\mathscr{W}}$ | 7.1336 | 0.007565 |
| $\xi_{\mathscr{R}}$ | 14.405 | 0.0001474 |
| $\xi^*_{\mathscr{R}}$ | 14.403 | 0.0001476 |
| $\xi_{\mathscr{T}}$ | 9.1896 | 0.002434 |

We note for this example that all tests would reject the null hypothesis $\mathscr{H}_0 : \beta_4 = 0$ for any the usual significance levels of 0.10, 0.05 and 0.01, despite the numerical difference in the estimated values. Also, the gradient test showed an estimated value closed to the likelihood ratio test and therefore a very similar $p$ value.

### 5.2.2 Air Sampler Data

Friedl & Stadlober (1997) and Friedl (2013) describes a data from from environmental microbiology study. In this study, airborne micro-organisms were monitored at seven outdoor sample sites in the adjacencies of Graz, Austria. The sample collection ran every two weeks during a period of a year. For our analysis, we consider the

subsample of two sites (`site = 6` and `site = 7`). This study was performed using a six stage Andersen air sampler which collected particles (also known as bioaerosols or biological aerosols) at a rate of $\approx 28.31$ litres (1 $ft^3$) per minute. Each stage of the air sample contains a Petri dish with a proper agar medium where the micro-organisms may be found. The sampler ran for four minutes then each of the Petri dish was removed and after incubation, the number of colonies formed units (`cfu`) was counted.

It has been registered then the $\mathtt{b}_j$ and $\mathtt{f}_j$ data, $j = 1, ..., 6$ stages, which provides information on the number of `cfu`'s observed in 128.3 litres of air for bacteria and fungi, respectively. Thus, we have the following variables.

`date` date when measurement was done (in format dd.mm)

`site` indicates the site where the measures were collected (1–7)

`humi` relative humidity in percent

`temp` temperature in degree Celsius

$\mathtt{b}_j$ bacteria `cfu`'s sampled on stage $j$, for $j = 1, ..., 6$

$\mathtt{f}_j$ fungi `cfu`'s sampled on stage $j$, for $j = 1, ..., 6$

Some observations were excluded: from 1995 May, 16 – site 3, October, 17 – site 1, November, 28 – site 4; and from 1996 January, 3 – site 5 because of some measurement error. Because of that, the total sample size of the dataset is 178 ($7 \times 26 - 4$). For our analysis, we consider the subset corresponding to the stages 5 and 6 and sites 6 and 7 only giving a sample size of 150 observations. We can read the full dataset and take the subset for our analysis by using the code below.

```
bacteria = read.table("http://www.stat.tugraz.at/courses/files/
                       bacteria.dat",
                      head = TRUE)
head(bacteria)

##   date site humi temp b1 b2 b3 b4 b5 b6 f1 f2 f3 f4 f5 f6
```

```
## 1 8.03    1   31   13 17 17 19 10 17 21 10  4  1  0  4  1
## 2 8.03    2   32    9 10  6  2 16  7  2 12  6  7  4  2  2
## 3 8.03    3   28    8  5  0  3  2  8  1  7  1  0  3  4  0
## 4 8.03    4   28    9  2  1  4  4  4  0 10  3  3 16  6  0
## 5 8.03    5   28   11  8 11 11  2  3  8  2  1  4 13 10  3
## 6 8.03    6   29   10  2  1  0  2  1  1  0  1  0  2  2  0

bac       <- bacteria[(bacteria$site > 5), ]
b.total   <- bac$b4 + bac$b5 + bac$b6
date.crit <- bac$date[b.total > 20]
bac       <- bac[(bac$date != date.crit), ]


var.sel <- c("date", "site", "humi", "temp")
bac <- rbind(cbind(bac[ ,var.sel], stage = 4, cfu = bac$b4),
             cbind(bac[ ,var.sel], stage = 5, cfu = bac$b5),
             cbind(bac[ ,var.sel], stage = 6, cfu = bac$b6))
bac$date  <- factor(bac$date)
bac$site  <- factor(bac$site)
bac$stage <- factor(bac$stage)
```

Let `cfu` be the response variable and consider the case that

$$\texttt{cfu}_i \overset{\text{ind}}{\sim} \mathcal{P}oi(\lambda = \mu_{ik}) \quad \text{for} \quad i = 1, \ldots, 150 \quad k = 1, 2,$$

where $\mu_{ik} = \mathrm{E}[\texttt{cfu}_i | z_k]$ which is linked to the explanatory variables by

$$\log(\mu_{ik}) = \beta_1 \texttt{stage}_{5i} + \beta_2 \texttt{stage}_{6i} + \beta_3 \texttt{site}_{7i} + \beta_4 (\texttt{stage}_{5i}\texttt{site}_{7i}) +$$
$$+ \beta_5 (\texttt{stage}_{6i}\texttt{site}_{7i}) + \beta_6 \texttt{temp}_i + \beta_7 \texttt{temp}_i^2 + z_k,$$

where $\beta_j$ for $j = 1, \ldots, 7$ are unknown parameters and $z_k$ is an unobserved random effect, for $k = 1, 2$. Our approach here supposes that the distribution of $z_k$ is unspecified which allow us to use the NPML estimation method. Therefore, one can use the code below to fit this model.

```
fit <- alldist(cfu ~ stage*site + temp + I(temp^2),
               data = bac, family = poisson,  k = 2,
               random.distribution = "np",
               plot.opt = 0, verbose = FALSE)
summary(fit)

##
## Call:  alldist(formula = cfu ~ stage * site + temp + I(temp^2),
##                family = poisson, data = bac, k = 2,
##                random.distribution = "np", plot.opt = 0,
##                verbose = FALSE)
##
## Coefficients:
##                  Estimate   Std. Error     t value
## stage5        0.394199122 0.1972243502  1.99873455
## stage6       -0.256549988 0.2337363398 -1.09760420
## site7         0.145078383 0.2124451177  0.68289817
## temp          0.058464594 0.0175809700  3.32544758
## I(temp^2)    -0.002053258 0.0006214078 -3.30420297
## stage5:site7 -0.505734904 0.2888280442 -1.75098961
## stage6:site7  0.238191454 0.3100037225  0.76835030
## MASS1         0.019053172 0.1913314231  0.09958203
## MASS2         1.325236225 0.2009559641  6.59465983
##
## Mixture proportions:
##    MASS1     MASS2
## 0.870212  0.129788
##
## Random effect distribution - standard deviation:    0.438969
##
## -2 log L:     526.5     Convergence at iteration  56
```

We wish to test

$$\begin{cases} \boldsymbol{\beta}_1 = \mathbf{0} \\ \boldsymbol{\beta}_1 \neq \mathbf{0} \end{cases},$$

where $\boldsymbol{\beta}_1 = (\beta_6, \beta_7)^\top$ from $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ with $\boldsymbol{\beta}_2 = (\beta_1, \ldots, \beta_5)^\top$. This means that we are testing if `temp` plus its quadratic effect has some impact in the model. We have therefore to fit the model without this effect first, which can be done via the R code below.

```
fit.null <- alldist(cfu ~ stage*site, data = bac,
                    family = poisson,  k = 2,
                    random.distribution = "np",
                    plot.opt = 0, verbose = FALSE)
summary(fit.null)

##
## Call:  alldist(formula = cfu ~ stage * site, family = poisson,
##                data = bac, k = 2, random.distribution = "np",
##                plot.opt = 0, verbose = FALSE)
##
## Coefficients:
##                Estimate Std. Error    t value
## stage5        0.4160817  0.1971942  2.1100096
## stage6       -0.2816875  0.2334704 -1.2065238
## site7         0.1360769  0.2122844  0.6410123
## stage5:site7 -0.4919003  0.2887629 -1.7034746
## stage6:site7  0.2709082  0.3094782  0.8753710
## MASS1         0.1989127  0.1599253  1.2437853
## MASS2         1.5629860  0.1700300  9.1924149
##
## Mixture proportions:
##     MASS1      MASS2
## 0.8651551  0.1348449
```

```
##
## Random effect distribution - standard deviation:    0.4659099
##
## -2 log L:     534.2     Convergence at iteration  52
```

Then, the tests can be computed using the following code.

```
lr.test(fit.null,~temp+I(temp^2))

##
##  Likelihood ratio test for GLMwRE
##
##  null model: cfu ~ stage * site
##  alt. model: cfu ~ stage + site + temp + I(temp^2) + stage:site
##
## statistic = 7.7185, parameter = 2, p-value = 0.02108

wald.test(fit.null,~temp+I(temp^2))

##
##  Wald test for GLMwRE
##
##  null model: cfu ~ stage * site
##  alt. model: cfu ~ stage + site + temp + I(temp^2) + stage:site
##
## statistic = 14.286, parameter = 2, p-value = 0.0007903

rao.test(fit.null,~temp+I(temp^2))

##
##  Rao test for GLMwRE
##
##  null model: cfu ~ stage * site
##  alt. model: cfu ~ stage + site + temp + I(temp^2) + stage:site
```

```
##
## statistic = 3.6195, parameter = 2, p-value = 0.1637

gradient.test(fit.null,~temp+I(temp^2))

##
##  Gradient test for GLMwRE
##
##  null model: cfu ~ stage * site
##  alt. model: cfu ~ stage + site + temp + I(temp^2) + stage:site
##
## statistic = 7.7872, parameter = 2, p-value = 0.02037
```

For this example, we observe that the Wald statistic has higher value ($\hat{\xi}_W = 14.286$, $p$-value $\approx 0.0007903$) and the Rao statistic has the lowest value ($\hat{\xi}_R = 3.6195$, $p$-value $\approx 0.1637$). We have then that, for any of the usual significance levels (10%, 5% and 1%) that the Wald test would reject $\mathscr{H}_0$ and the Rao test would not reject $\mathscr{H}_0$. However, the likelihood ratio and gradient statistics have similar numbers ($\hat{\xi}_{\mathscr{LR}} = 7.7185$ and $\hat{\xi}_{\mathscr{T}} = 7.7872$, respectively) and consequently $p$-values $\approx 0.02$.

### 5.2.3 Gene sequencing data

The data in this application comprise the results of a gene sequencing study from Elsensohn et al. (2017). The study evaluates the performance of two pipelines, an academic (BWA-GATK) and a commercial (TMAP-NextGENe), in terms of the number of chromosomal positions identified as non-variants on a panel of 41 genes in 43 epileptic patients.

This data set contains the following variables:

`Ident` : patient identification,

`Nbtot` : number of chromosomal positions identified as non-variants,

`NbVarTotal` : number of total chromosomal positions,

`BG` : factor which represents the variant identified by pipeline BWA-GATK, with two levels: (`BG=1`) or not (`BG=0`),

`NG` : factor with two levels representing the variant identified by pipeline TMAP-NextGENe (`NG=1`) or not (`BG=0`),

`Common` : factor with two levels that indicate if the variants are found in both pipelines (`Common=1`) or not (`Common=0`),

`Nature` : factor for variant "identity".

```
dt <- load(url("https://static-content.springer.com/esm/art%3A10.1186
                %2Fs12859-017-1552-9/MediaObjects/12859_2017_1552_MOE
                SM2_ESM.rdata"))
save(dt, file = "Additinal File 2.RData")
source(url("https://static-content.springer.com/esm/art%3A10.1186%2Fs
            12859-017-1552-9/MediaObjects/12859_2017_1552_MOESM1_ESM.
            r"),
       echo = FALSE)
pipelines <- TableContinVar5cell

pipelines$BG      <- as.factor(pipelines$BG)
pipelines$NG      <- as.factor(pipelines$NG)
pipelines$Common  <- as.factor(pipelines$Common)
pipelines$MargeEq <- as.factor(pipelines$MargeEq)
pipelines$Nature  <- as.factor(pipelines$Nature)
pipelines$prop    <- with(pipelines,Nbtot/NbVarTotal)
```

Let $\pi_{ijk} = \mathrm{E}[\texttt{Nbtot}_{ijk}/\texttt{NbVarTotal}_{ijk}|z_k, u_k]$ the expected proportion of chromosomal positions identified as non-variants for the $i$th subject, $i = 1, \ldots, 43$, $j$th,

$j = 1, \ldots, 5$ replicate and $k$th mass, $k = 1, \ldots, 4$. We assume the linear predictor

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \beta_1 \texttt{BG}_{ij} + \beta_2 \texttt{NG}_{ij} + \beta_3 \texttt{Common}_{ij} + \beta_4 \texttt{Nature}_{ij} + \\ + z_k + u_k \texttt{Common}_{ij}, \quad i = 1, \ldots, 43, \quad j = 1, \ldots, 5, \quad k = 1, \ldots, 4, \tag{5.2.11}$$

where $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$ are the unknown fixed effects parameters, $z_k$ represents the random intercepts and $u_k$ the random slopes for the factor `Common`. Figure 5.41 show the EM trajectories for the computed disparity ($-2\ell$) and for the mass points. See below the code used to fit the model in (5.2.11).

```
fit <- allvc(prop ~ BG + NG + Common + Nature,
             random = ~ Common|Ident,
             k = 4, weights = NbVarTotal,
             data = pipelines, family = binomial(logit),
             tol = .2, plot.opt = 0, verbose = FALSE)
summary(fit)

##
## Call:  allvc(formula = prop ~ BG + NG + Common + Nature,
##              random = ~Common|Ident, family = binomial(logit),
##              data = pipelines, k = 4, tol = 0.2,
##              weights = NbVarTotal, plot.opt = 0,
##              verbose = FALSE)
##
## Coefficients:
##                    Estimate  Std. Error       t value
## BG1           -11.099144026 0.006214437 -1786.0257241
## NG1           -10.666739180 0.005279753 -2020.3101846
## Common1        10.998166559 0.011867248   926.7663846
## Nature1        -1.436302250 0.010429469  -137.7157618
## MASS1           4.725709922 0.004675862  1010.6607766
```

```
## MASS2           4.804489754 0.004664099  1030.1003190
## MASS3           4.875188484 0.004431149  1100.2085860
## MASS4           5.002672115 0.005006465   999.2424510
## MASS1:Common1   0.069787287 0.013738177     5.0798069
## MASS2:Common1   0.162814753 0.013096735    12.4317051
## MASS3:Common1  -0.007777029 0.012834039    -0.6059689
##
## Mixture proportions:
##     MASS1      MASS2      MASS3      MASS4
## 0.2325690  0.2453434  0.2895360  0.2325516
##
## Random effect distribution - standard deviation:    0.09847663
##
## -2 log L:     5505.3     Convergence at iteration  12
```

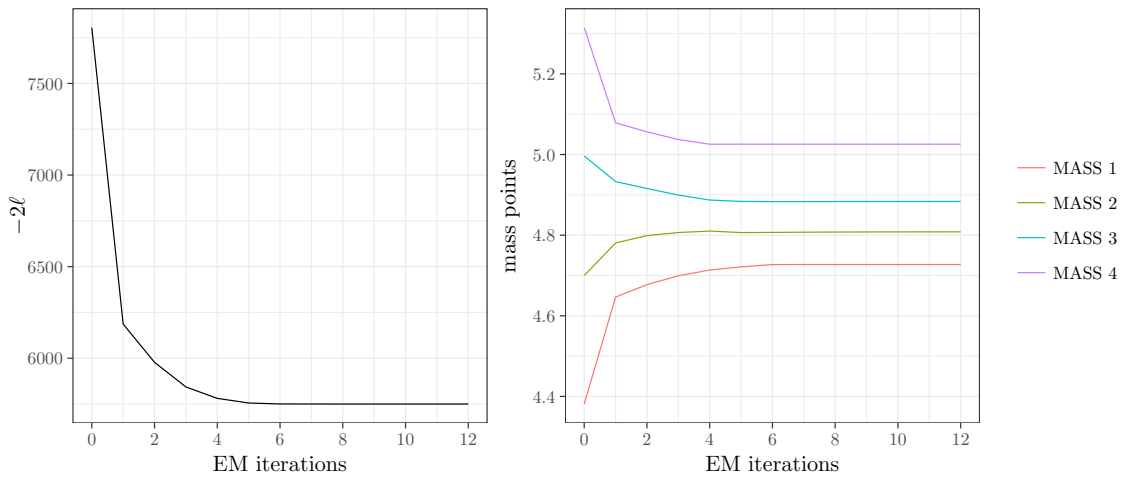


Figure 5.41: disparity values over iterations (left) and mass points estimates over iterations (right) for the model in (5.2.11) fitted using `allvc`.

We might be interested in testing if the effect of both pipelines interfere on the proportion of chromosome variants identified. This idea can be translated into the

hypothesis

$$\begin{cases} \mathscr{H}_0 : \boldsymbol{\beta}_1 = \mathbf{0} \\ \mathscr{H}_1 : \boldsymbol{\beta}_1 \neq \mathbf{0} \end{cases},$$

where $\boldsymbol{\beta}_1 = (\beta_1, \beta_2)^\top$, a partition from the vector of parameters $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ with $\boldsymbol{\beta}_2 = (\beta_3, \beta_4)^\top$. In R, we have to fit the model under the null hypothesis and then run the likelihood ratio and gradient tests, which can be seen in the piece of code below.

```
#####
# testing for 'BG' and 'NG'
#####
fit.null <- allvc(prop ~ Common + Nature, random = ~ Common|Ident,
                  k = 4, weights = NbVarTotal,
                  data = pipelines, family = binomial(logit),
                  tol = .2, plot.opt = 0, verbose = FALSE)
lr.test(fit.null, ~ BG + NG)

##
##  Likelihood ratio test for GLMwRE
##
##  null model: prop ~ Common + Nature
##  alt. model: prop ~ Common + Nature + BG + NG
##
## statistic = 61352000, parameter = 2, p-value < 2.2e-16

gradient.test(fit.null, ~ BG + NG)

##
##  Gradient test for GLMwRE
##
##  null model: prop ~ Common + Nature
##  alt. model: prop ~ Common + Nature + BG + NG
##
## statistic = 308.79, parameter = 2, p-value < 2.2e-16
```

We observe that the likelihood ratio statistic is very high ($\hat{\xi}_{\mathscr{LR}} = 61352000$) which indicates that it is quite likely that $\mathscr{H}_0$ is not true based on this sample ($p$-value $\approx 0$). The gradient test statistic has much smaller value ($\hat{\xi}_{\mathscr{T}} = 308.79$) however still implies rejection of $\mathscr{H}_0$ ($p$-value $\approx 0$). Therefore, according to the tests, both `BG` and `NG` should remain in the model.

### 5.2.4 Redness data

We take the data from an experiment given by Markussen (2017). It is of interest to investigate how the continuous measurement of redness of pork meat after slaughter is affected by the storage (in light or darkness), by the time (1, 4 or 6 days) and by the breed (old and new, 10 pigs each). Six chops were taken from each pig and allocated according to the scheme shown in Table 5.13. This gives $2 \times 10 \times 6 = 120$

Table 5.13: Factor allocation [source: Markussen (2017)].

| Storage | 1 days | 4 days | 6 days |
|---|---|---|---|
| Dark | chop 1 | chop 2 | chop 3 |
| Light | chop 4 | chop 5 | chop 6 |

samples of pork chops in total. Given that the vector of response variables $\mathbf{y}$ is strictly positive, we consider that the redness measurements of a given replicate corresponding to the $i$th breed, the $j$th storage and the $k$th time are independently distributed as inverse Gaussian with means $\mu_{ijk}|Z$ and a fixed dispersion parameter. We also assume the linear predictor is linked to $\mu_{ijk}$ as

$$\mu_{ijk}|Z = Z + \alpha_i + \tau_j + \beta_k \quad i = 1, 2, \quad j = 1, 2, \quad k = 1, 2, 3 \tag{5.2.12}$$

where $Z$ is a random intercept representing the base level for each pig, $\alpha_1 = \tau_1 = \beta_1 = 0$ and (5.2.12) is one of the configurations of the variance component model defined in Aitkin et al. (2009). Because $Z$ is an unknown random variable, the EM approach in conjunction with the maximum likelihood method can be applied for parameter estimation. We assume that the distribution of $Z$ is unspecified for all the model adjustments and, for estimation purposes we used the Nonparametric maximum likelihood (NPML) of Einbeck & Hinde (2006a).

### The gradient test

Consider including in 5.2.12 the interaction between storage and time, i.e. testing the null hypothesis $\mathscr{H}_0 : ((\tau\beta)_{22}, (\tau\beta)_{23})^\top = 0$. Let $\ell$ be the total log-likelihood and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ the vector of fixed effects parameters where $\boldsymbol{\theta}_1 = ((\tau\beta)_{22}, (\tau\beta)_{23})^\top$ is our vector of parameters of interest and $\boldsymbol{\theta}_2$ is a vector of nuisance parameters. The unrestricted MLE for $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^\top, \hat{\boldsymbol{\theta}}_2^\top)^\top$ and the restricted to the null hypothesis is $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1^{0\top}, \tilde{\boldsymbol{\theta}}_2^\top)^\top$, where $\boldsymbol{\theta}_1^{0\top}$ is an arbitrary vector (in our application is equal to 0, for instance). From now on the top accents $\wedge$ and $\sim$ represent the MLE unrestricted and restricted to the null hypothesis. Let $\mathscr{U} = \partial\ell/\partial\boldsymbol{\theta} = (\mathscr{U}_1^\top, \mathscr{U}_2^\top)^\top$ the respective partitioned score vector. Terrell (2002) proposed the gradient statistic for testing $\mathscr{H}_0$ denoted as $\xi_{\mathscr{T}} = \tilde{\mathscr{U}}_1^\top(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^0)$. Note that $\xi_{\mathscr{T}}$ does not have any matrix computation in its formula which turns to be its main advantage. In theory, the reference distribution for $\xi_{\mathscr{T}}$ is $\chi^2_q$ where $q$ denote the dimension of $\boldsymbol{\theta}_1$. Because of that, $\xi_{\mathscr{T}}$ is comparable to the $\xi_{\mathscr{LR}}$, the likelihood ratio statistic. Table 5.14 shows the estimates for the test statistics, the chi-squared $p$ values and the equivalent bootstrap version.

Table 5.14: Likelihood ratio and gradient tests for the null hypothesis. The $p$ values were computed using the chi-square distribution with two degrees of freedom and * empirical bootstrap as the reference distributions.

| | likelihood ratio | gradient |
|---|---|---|
| Statistic | 8.794883 | 10.25232 |
| $p$ value | 0.01230879 | 0.005939328 |
| $p$ value* | 0.01880188 | 0.00730073 |

### Bootstrap and confidence intervals

The main purpose of the bootstrap experiment here is to verify how accurate is the chi-square approximation for the test statistics. We propose therefore a bootstrap in two levels taking the model under null hypothesis as true. In the first level we resample the estimated random intercepts obeying the respective estimated probabilities (nonparametric bootstrap) and in the second level we generate responses given the new intercepts (parametric bootstrap). Then the model in (5.2.12) is fit-

ted and both likelihood ratio and gradient statistics are computed. We replicate the procedure 9999 times and the results can be seen in Figure 5.42. We also investigate

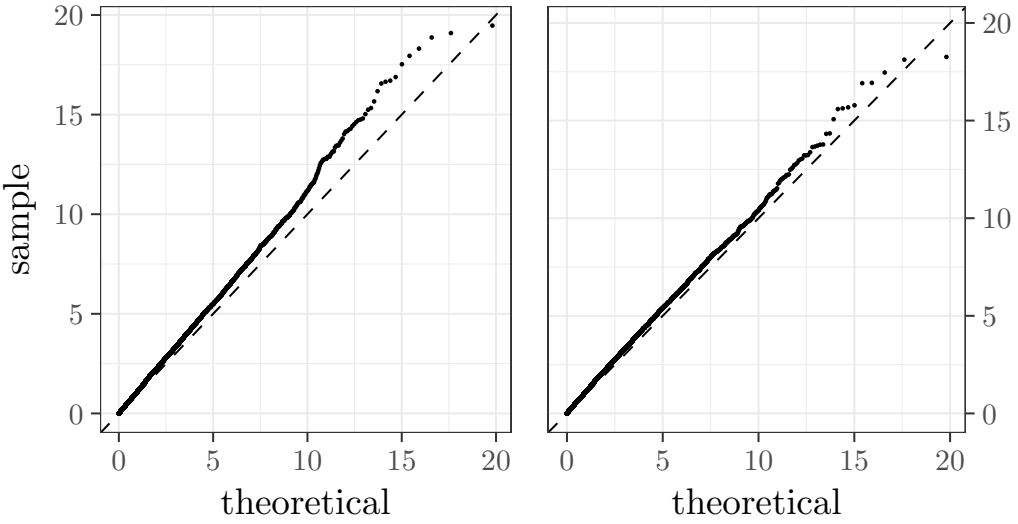


Figure 5.42: Bootstrap samples of the likelihood ratio statistic (left) and gradient statistic (right) compared to the theoretical $\chi^2_2$ for the test with hypothesis $\mathscr{H}_0$ : $(\tau\beta)_{22} = (\tau\beta)_{23} = 0$.

the power of the tests using the rejection rates under the alternative hypothesis $\mathscr{H}_1 : ((\tau\beta)_{22}, (\tau\beta)_{23})^\top = \delta\hat{\text{se}}((\widehat{(\tau\beta)}_{22}, \widehat{(\tau\beta)}_{23})^\top)$ with $\delta$ being a numeric sequence of 51 evenly spaced values in $[-3, 3]$. For each $\delta$ we generate 9999 bootstrap samples of $\xi_{\mathscr{LR}}$ and $\xi_{\mathscr{T}}$.

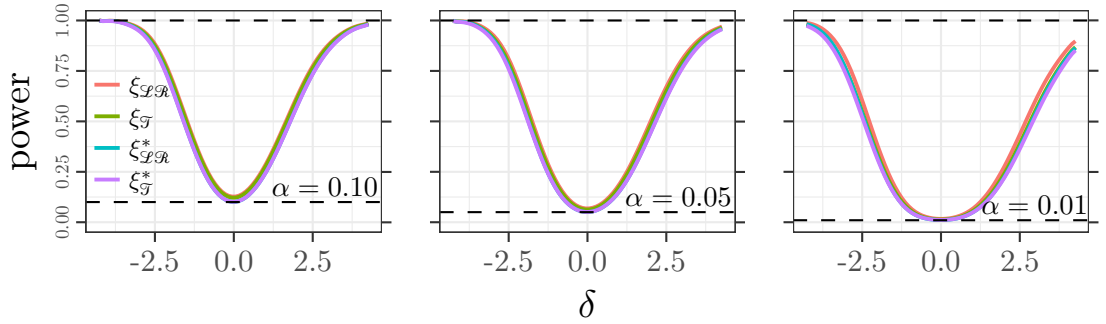


Figure 5.43: Bootstrap power of the likelihood ratio test and the gradient test for nominal levels of 10% (left), 5% (center) and 1% (right).

Figure 5.43 shows the estimated power curves where the two coloured lines — $\xi_{\mathscr{LR}}$ and — $\xi_{\mathscr{T}}$ represent the rejection rates for the likelihood ratio and gradient tests

taking $\chi^2_2$ as reference, respectively, and the coloured lines — $\xi^*_{\mathscr{LR}}$ and — $\xi^*_{\mathscr{T}}$ represent the likelihood and gradient tests taking the bootstrap distribution under $\mathscr{H}_0$ of each statistic as reference, respectively for $\alpha = 0.1$, 0.05 and 0.01. We note that the difference between the four curves is negligible.

We can produce confidence regions for $\boldsymbol{\theta}_1 = ((\tau\beta)_{22}, (\tau\beta)_{23})^\top$ inverting the gradient test however there is no analytic procedure so far. Numericaly, we took a grid of two sequences of 51 values for each $(\tau\beta)_{jk}$ on the interval $(\widehat{\tau\beta})_{jk} \pm 3\text{se}((\widehat{\tau\beta})_{jk})$. Then, we fit the model in (5.2.12) with $(\tau\beta)_{jk}$ as offset for each position of the grid and compute the test statistics for $\mathscr{H}_0$. Therefore, the region consists on the values of $\boldsymbol{\theta}_1 = ((\tau\beta)_{22}, (\tau\beta)_{23})^\top$ that satisfy $\xi_{\mathscr{T}} < \chi^2_2$. The same procedure has been done for $\xi_{\mathscr{LR}}$. The Figure 5.44 shows the contour maps for the 90% confidence regions.

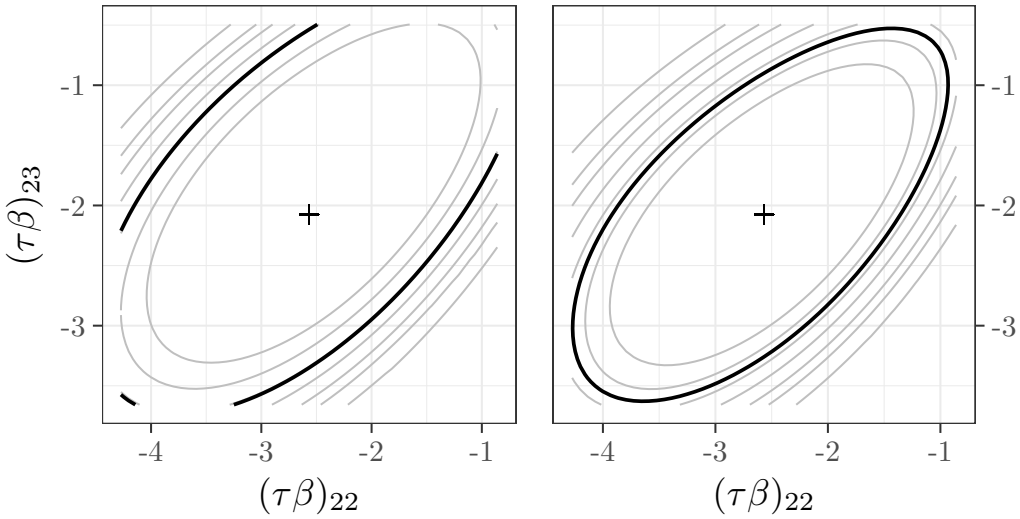


Figure 5.44: 90% confidence regions in black for $(\tau\beta)_{22}$ and $(\tau\beta)_{23}$ based on the numerical inversion of the likelihood ratio test (left) and the gradient test (right).

# Chapter 6

# Conclusion

The gradient test is a useful and important asymptotic test such as the likelihood ratio, Wald and Rao tests. The gradient statistic is computationally less expensive than the Wald and Rao statistics because it does not have any matrices or matrix operations in its formula. This turns to be one of the gradient test most appealing features. The properties of gradient test have been studied for several different types of models since its conception. However, before this thesis, the performance of the gradient test had not been assessed in the context of random effects modelling. In this thesis, we have argued that the gradient test is a solid alternative to the likelihood ratio, Wald and Rao tests for GLMwRE considering the type I error and the power for finite samples.

Central to this work is the development of the formulae and notation required to compute the gradient statistic for GLMwRE. We proposed in this thesis a comprehensive matrix notation for the GLMwRE and for the score vector and the Fisher information matrix. Despite the fact that the Fisher information matrix is not required for the gradient statistic, we made this endeavour to obtain the Wald and Rao statistic formulae. The GLMwRE definition, notation and gradient statistic are defined in Chapter 3 and the Fisher information for GLMwRE in Chapter 4.

A significant focus of this work was to quantify two properties of the gradient test, the type I error and the power estimated by the rejection rates under null and alternative hypothesis, respectively, for different settings of the GLMwRE. To do this, we conducted an extensive simulation experiment which covers several different con-

figurations of the GLMwRE presented in Chapter 5. For simulation purposes, we grouped the models in three main classes, here named *Gaussian quadrature models*, non-parametric models and variance components models. The first two refers to the estimation process and the choice of random effects distribution together, normal random effects — Gaussian quadrature estimation and unspecified distribution — non-parametric maximum likelihood, respectively. These two are frequently applied to where the classic GLM cannot deal with overdispersion. The last class, the variance components model, is a generalisation of the overdispersion model for grouped data. For each of this classes, we explored the gradient test for testing parameters regarding to the fixed effects and for different possible response distributions and for a range of different sample sizes. In parallel, we compared the gradient test to the likelihood ratio, Wald and Rao tests for the same scenarios.

Based on the simulation results presented in Section 5.1, it can be concluded that the gradient test is preferred over the classic likelihood ratio, Wald and Rao tests. A few points must be stressed here about the behaviour of the test. All four test statistics have asymptotically chi-square distribution. However, for a finite sample size, some difference between the distribution of the test statistic and the chi-square distribution is to be expected. This translates to some difference between the rejection rates and the true nominal levels. For smaller sample sizes, we noticed that rejection numbers of the four tests are far from the nominal levels with some advantage for the gradient test. This behaviour intensifies for the models estimated with Non-parametric maximum likelihood, both NP and VC models. We also observed that the difference to the nominal level increases as the number of mass points increase for these models. We did not see this for GQ models. In all cases the numbers improve as the sample size increase. Overall, we observed that the rejection rates of the gradient statistic are fairly close to the true nominal level in all scenarios. The likelihood ratio test is the second best followed by the Rao test and the Wald test showed the worst numbers.

The power simulations were performed under the same conditions as the type I error simulations. In fact, there is a trade-off between the high nominal levels and the estimated power which leads to artificially higher values for power. This phenomena

is clearly seen in the results where the Wald and Rao tests showed very high curves compared to the other two. On the other side, the gradient test showed power curves not very distant from the ones produced by the likelihood ratio test despite the fact that the gradient test showed better approximation to the nominal levels under the same conditions.

In summary, the message is that the gradient test overall outperformed the well established asymptotic tests in terms of type I error without much power loss for GLMwRE. This advantage plus the fact that the gradient test has statistic computationally less costly than the Wald and Rao statistics support the idea that the gradient test should be preferred in the context of GLMwRE.

# Bibliography


Abramowitz, M. & Stegun, I. A. (1972). *Handbook of mathematical functions with formulas, graphs and mathematical tables*. Washington: US Government Printing Office. 10th printing.

Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.

Aitkin, M. (1994). An em algorithm for overdispersion in generalised linear models. In *Proceedings of the 9th International Workshop on Statistical Modelling*. 1–8.

Aitkin, M. (1996a). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* 6 251–262.

Aitkin, M. (1996b). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and computing* 6 251–262.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* 55 117–128.

Aitkin, M. & Francis, B. J. (1995). Fitting overdispersed generalized linear models by non-parametric maximum likelihood. *GLIM newsletter* 25 37–45.

Aitkin, M. A., Francis, B., Hinde, J. & Darnell, R. (2009). *Statistical modelling in R*. Oxford University Press Oxford.

Anderson, D. A. (1988). Some models for overdispersed binomial data. *Australian & New Zealand Journal of Statistics* 30 125–148.

Anderson, D. A. & Hinde, J. P. (1988). Random effects in generalized linear models and the EM algorithm. *Communications in Statistics-Theory and Methods* 17 3847–3856.

Bates, D. & Maechler, M. (2017). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-10, URL `https://CRAN.R-project.org/package=Matrix`.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, URL `http://CRAN.R-project.org/package=lme4`.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

Chen, J. & Li, P. (2009). Hypothesis test for normal mixture models: The em approach. *The Annals of Statistics* 2523–2542.

Cordeiro, G. M. (1999). *Introdução a teoria assintótica*. Rio de Janeiro: IMPA.

Cramér, H. (1999). *Mathematical methods of statistics*, vol. 9. Princeton university press.

da Silva-Júnior, A. H. M. (2017). Gradient test for variance component models. In M. Grzegorczyk & G. Ceoldo, eds., *Proceedings of the 32nd International Workshop on Statistical Modelling*, vol. 2. 71–74.

da Silva-Júnior, A. H. M., da Silva, D. N. & Ferrari, S. L. P. (2014). mdscore: An R package to compute improved score tests in generalized linear models. *Journal of Statistical Software* 61 1–16. URL `http://www.jstatsoft.org/v61/c02/`.

da Silva-Júnior, A. H. M., Einbeck, J. & Craig, P. S. (2015). The gradient test for generalised linear models with random effects. In A. Blanco-Fernandez &

G. Gonzalez-Rodriguez, eds., *Programme and Abstracts: 9th International Conference on Computational and Financial Econometrics (CFE 2015) and 8th International Conference of the ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on Computing & Statistics (ERCIM 2015)*. 63.

da Silva-Júnior, A. H. M., Einbeck, J. & Craig, P. S. (2016). Gradient test on generalised linear models with random effects. In J.-F. Dupuy & J. Josse, eds., *Proceedings of the 31st International Workshop on Statistical Modelling*, vol. 1. 213–218.

da Silva-Júnior, A. H. M., Einbeck, J. & Craig, P. S. (2017). Fisher information on Gaussian quadrature models. *Statistica Neerlandica* In press.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)* 1–38.

Einbeck, J., Darnell, R. & Hinde, J. (2014). *npmlreg: Nonparametric maximum likelihood estimation for random effect models*. R package version 0.46-1, URL `https://CRAN.R-project.org/package=npmlreg`.

Einbeck, J. & Hinde, J. (2006a). A note on NPML estimation for exponential family regression models with unspecified dispersion parameter. *Austrian journal of statistics*. 35 233–243.

Einbeck, J. & Hinde, J. (2006b). Random effect modelling for regression models with gamma-distributed response. Central Statistics Office (Ireland).

Elsensohn, M., Leblay, N., Dimassi, S., Campan-Fournier, A., Labalme, A., Roucher-Boulez, F., Sanlaville, D., Lesca, G., Bardel, C. & Roy, P. (2017). Statistical method to compare massive parallel sequencing pipelines. *BMC Bioinformatics* 18 139.

Ferrari, S. & Pinheiro, E. C. (2014). Small-sample likelihood inference in

extreme-value regression models. *Journal of Statistical Computation and Simulation* 84 582–595.

Friedl, H. (2013). Generalisierte lineare modelle.

Friedl, H. & Stadlober, E. (1997). Resampling methods in generalized linear models useful in environmetrics. *Environmetrics* 8 441–457.

Golub, G. H. & Welsch, J. H. (1969). Calculation of gauss quadrature rules. *Mathematics of Computation* 23 221–230.

Hayakawa, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika* 62 451–460.

Heinze, G. & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in medicine* 21 2409–2419.

Hinde, J. (1982). Compound poisson regression models. In *GLIM 82: Proceedings of the International Conference on Generalised Linear Models*. Springer, 109–121.

Kosmidis, I. (2017). *brglm2: Bias reduction in generalized linear models*. R package version 0.1.4, URL `https://github.com/ikosmidis/brglm2`.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* 73 805–811.

Lehmann, E. L. & Casella, G. (1998). *Theory of point estimation*, vol. 31. Springer Science & Business Media.

Lemonte, A. (2016). *The Gradient Test: Another Likelihood-Based Test*. Academic Press.

Lemonte, A. J. (2012). Local power properties of some asymptotic tests in symmetric linear regression models. *Journal of Statistical Planning and Inference* 142 1178–1188.

Lemonte, A. J. (2013). Nonnull asymptotic distributions of the lr, wald, score and gradient statistics in generalized linear models with dispersion covariates. *Statistics* 47 1249–1265.

Lemonte, A. J. & Ferrari, S. L. (2011a). Local power of the lr, wald, score and gradient tests in dispersion models. *arXiv preprint arXiv:1102.4371* .

Lemonte, A. J. & Ferrari, S. L. (2011b). Size and power properties of some tests in the birnbaum–saunders regression model. *Computational Statistics & Data Analysis* 55 1109–1117.

Lemonte, A. J. & Ferrari, S. L. (2011c). Testing hypotheses in the birnbaum–saunders distribution under type-ii censored samples. *Computational Statistics & Data Analysis* 55 2388–2399.

Lemonte, A. J. & Ferrari, S. L. (2012). Local power and size properties of the lr, wald, score and gradient tests in dispersion models. *Statistical Methodology* 9 537–554.

Lemonte, A. J., Ferrari, S. L. et al. (2012). A note on the local power of the lr, wald, score and gradient tests. *Electronic Journal of Statistics* 6 421–434.

Liang, K.-Y. (1984). The asymptotic efficiency of conditional likelihood methods. *Biometrika* 71 305–313.

Lindsay, B., Clogg, C. C. & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association* 86 96–107.

Lindsay, B. G. (1995). Mixture models: theory, geometry and applications. In *NSF-CBMS regional conference series in probability and statistics*. JSTOR, i–163.

Markussen, B. (2017). Lecture notes for statistical methods for the biosciences.

Medeiros, F. M., da Silva-Júnior, A. H., Valença, D. M. & Ferrari, S. L. (2014). Testing inference in accelerated failure time models. *International Journal of Statistics and Probability* 3 121–131.

Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* 135 370–384.

Ostle, B. & Mensing, R. W. (1963). *Statistics in research*. Ames: Iowa State University Press.

Peers, H. (1971). Likelihood ratio and associated test criteria. *Biometrika* 58 577–587.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Proceedings of the Cambridge Philosophical Society*, vol. 44. Cambridge University Press, 50–57.

Shuster, J. J. & Miura, C. (1972). Two-way analysis of reciprocals. *Biometrika* 59 478–481.

Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics* 34 206–215.

Vargas, T. M., Ferrari, S. L. & Lemonte, A. J. (2014). Improved likelihood inference in generalized linear models. *Computational Statistics & Data Analysis* .

Vargas, T. M., Ferrari, S. L., Lemonte, A. J. et al. (2013). Gradient statistic: Higher-order asymptotics and bartlett-type correction. *Electronic Journal of Statistics* 7 43–61.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* 54 426–482.

Wilks, S. S. et al. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* 9 60–62.

# Appendix A

# R code

In this Appendix we describe the source code of R functions used in Chapter 5. The first function code is `respvar` in A.1 which computes the estimated response variance based on the formulae proposed in Section 4.2 for GLMwRE.
In the sequence we have source codes of the functions to compute the likelihood ratio test, `lr.test`, in Section A.2, the Wald test, `wald.test`, in Section A.3, the Rao test, `rao.test`, in Section A.4 and the gradient test, `gradient.test`, in Section A.5, respectively.
Finally, we have the function `print.test` in Section A.6 that automatically works on the background and it is responsible for the standard output of the four test functions in R.
All the functions use R object resultant of fitted models using either `alldist` or `allvc` from package **npmlreg** (Einbeck et al., 2014).

## A.1 Function to estimate the response variance

We present in this Section the code to estimate the response variance. This function has two arguments, `m` which corresponds to the model fitted by `alldist` or `allvc` and `exact` either be `TRUE` or `FALSE` which indicates if the estimation is using the analytic or the EM approximation. This is only valid for GQ models which means that for NPML models the `exact=FALSE` by default. Comments indicated after `#` on the code describe which piece computes in R.

```
respvar = function(m,exact=TRUE){
  # extract the 'phi' estimate according to the distribution.
  phi = switch(m$family$family,
               gaussian         = m$sdev$sdev^-2,
               Gamma            = m$shape$shape,
               inverse.gaussian = m$shape$shape,
               poisson          = 1,
               binomial         = 1
  )

  # if 'm' is a GQ model and 'exact=TRUE' the function will
  # compute the response otherwise will estimate using
  # the last EM results.
  if(class(m)=="glmmGQ"&exact==TRUE){
    # extract some quantities from the fitted model.
    p     = length(m$coefficients)  # no. of coefficients
    N     = length(m$y)             # stacked sample size
    K     = length(m$masses)        # number of mass points
    n     = N/K                     # sample size 'n'
    X     = model.matrix(m)[1:n,-p] # model matrix
    beta  = coef(m)[-p]             # estim. fixed effects
    sigma = m$rsdev                 # estimated 'sigma'
    eta   = as.numeric(X%*%beta)    # estimated 'eta'

    # Here we have we have the response variance formulae
    # implementation for each combination of response distribution
    # and link function.
    respvar = switch(m$family$family,
      gaussian = switch(m$family$link,
        identity = phi^-1 + sigma^2,
        log      = phi^-1 + exp(2*eta+sigma^2)*(exp(sigma^2)-1),
        inverse  = phi^-1 + eta^-4*sigma^2+8*eta^-
                   6*sigma^4+15*eta^-8*sigma^6
        ),
      Gamma = switch(m$family$link,
        inverse  = phi^-1*(eta^-2+3*eta^-4*sigma^2)+
                   eta^-4*sigma^2+8*eta^-6*sigma^4+
```

```
38                  15*eta^-8*sigma^6,
39          identity = (phi^-1+1)*sigma^2 + phi^-1*eta^2,
40          log      = exp(2*eta+sigma^2)*((phi^-1+1)*exp(sigma^2)-1)
41          ),
42        inverse.gaussian = switch(m$family$link,
43          "1/mu^2" = phi^-1*(eta^-(3/2)+(15/8)*eta^-(7/2)*sigma^2)+
44                  (1/4)*eta^-3*sigma^2+(1/2)*eta^-5*sigma^4+
45                  (375/256)*eta^-7*eta^6,
46          inverse  = phi^-1*(eta^-3+6*eta^-5*sigma^2)+
47                  eta^-4*sigma^2+8*eta^-6*sigma^4+
48                  15*eta^-8*sigma^6,
49          identity = phi^-1*(eta^3+3*eta*sigma^2)+sigma^2,
50          log      = phi^-1*exp(3*eta+9*sigma^2/2)+
51                  exp(2*(eta+sigma^2))-
52                  exp(2*eta+sigma^2/2)
53          ),
54        poisson = switch(m$family$link,
55          log      = exp(eta+.5*sigma^2)+exp(2*(eta+sigma^2))-
56                  exp(2*eta+sigma^2),
57          identity = eta+sigma^2,
58          sqrt     = eta^2+4*eta^2*sigma^2+sigma^2+2*sigma^4
59          ),
60        binomial = switch(m$family$link,
61          logit    = exp(eta)/(exp(eta)+1)-exp(eta)^2/(exp(eta)+1)^2-
62                  (exp(eta)^2-exp(eta))*sigma^2/(2*exp(eta)+1)^3+
63                  (exp(eta)^3-exp(eta)^2)*sigma^2/(exp(eta)+1)^4-
64                  (exp(eta)^2-exp(eta))^2*sigma^4/
65                  (4*exp(eta)+1)^6,
66          probit   = pnorm(eta)-eta*sigma^2*dnorm(eta)/2-
67                  pnorm(eta)^2+eta*sigma^2*dnorm(eta)*pnorm(eta)-
68                  eta^2*sigma^4*dnorm(eta)^2/4,
69          cauchit  = 1/4-pi^-2*(atan(eta)-eta*sigma^2/(eta^2+1)^2)^2,
70          log      = exp(eta+sigma^2/2)-exp(2*eta+sigma^2),
71          cloglog  = exp(-exp(eta))-exp(-2*exp(eta))+
72                  (exp(2*eta)-exp(eta))*sigma^2/(2*exp(exp(eta)))-
73                  (exp(2*eta)-exp(eta))*sigma^2/exp(2*exp(eta))-
74                  (exp(4*eta)-2*exp(3*eta)+exp(2*eta))*sigma^4/
```

```
75                    (4*exp(2*exp(eta)))
76          )
77      )
78    } else{ # last EM response variance
79      K       = length(m$masses)                          # no. of masses
80      mu.est = m$family$linkinv(m$linear.predictors) # est. 'mu'
81      V.est   = m$family$variance(mu.est)                 # est. var. fun.
82      w       = as.vector(m$post.prob)                    # posterior prob
          .
83
84      # 'E[Var(y|z)]'
85      evz   = (phi^-1)*apply(w*matrix(V.est,byrow=FALSE,nc=K),1,sum)
86      # 'Var[E(y|z)]'
87      vmuz = apply(w*matrix(mu.est^2,byrow=FALSE,nc=K),1,sum)-
88        m$fitted.values^2
89
90      respvar = drop(evz + vmuz) # E[Var(y|z)] + Var[E(y|z)]
91    }
92    return(respvar) # the final result.
93  }
```

## A.2 Likelihood ratio test

The function `lr.test` computes the likelihood ratio test for GLMwRE and takes the arguments listed below.

`fit.null` the model fitted using either `alldist` or `allvc` under null hypothesis.

`subset.formula` this is the subset part of the linear predictor under alternative hypothesis which does not include the null hypothesis linear predictor. This should be informed as R `formula`.

`null.values` the values of $\boldsymbol{\beta}_1^0$ as a vector.

`sign.level` the significance level which by default is `0.05`.

More information is described embedded in the code as comments after `#`.

```
lr.test <- function(fit.null,
                    subset.formula,
                    null.values = 0,
                    sign.level = .05){
  # data name
  dname    <- fit.null$call$data
  # The full formula under alternative hypothesis
  frml     <- update(fit.null$formula,
                     formula(paste("~.+",subset.formula[-1])))
  # fit the model under alternative hypothesis
  fit.alt  <- update(fit.null,frml,plot.opt=0,verbose=FALSE)
  # some R code tricks to partition later the model matrix X
  cnmX     <- attr(fit.alt$coefficients,"names")
  cnmXnull <- attr(fit.null$coefficients,"names")
  vrtst    <- cnmX[!cnmX%in%cnmXnull]
  if(length(null.values)==1&length(vrtst)>1){
    null.values <- rep(null.values,length(vrtst))
  }
  attr(null.values, "names") <- vrtst
  # the full model matrix
  X        <- model.matrix(fit.alt)
  # the partition of the model matrix that interests us
  X1       <- cbind(X[,vrtst])

  if(any(null.values!=0)){
    fit.null <- update(fit.null,
                       offset=X1[1:nrow(fit.null$data),]%*%null.
                          values,
                       plot.opt=0,verbose=FALSE)
  }

  estimate   <- fit.alt$coefficients[vrtst]
  # the test statistic value
  slr        <- drop(fit.null$disparity - fit.alt$disparity)
  names(slr) <- "statistic"
  df         <- length(vrtst)
  names(df)  <- "parameter"
```

```
37    pval        <- pchisq(slr, df=df, lower=FALSE)
38    mthd        <- list("Likelihood ratio test for GLMwRE",
39                        "\n",
40                        paste("null model:",
41                              deparse(fit.null$formula)),
42                        paste("alt. model:", deparse(frml)))
43
44    rval <- list("statistic" = slr, "parameter" = df,
45                 "p.value" = pval, null.values = null.values,
46                 estimate = estimate, method = mthd,
47                 data.name = dname)
48
49    class(rval) <- "GLMwRE.test"
50    return(rval)
51  }
```

## A.3 Wald test

The `wald.test` function computes the Wald test for GLMwRE and takes the arguments listed below.

`fit.null` the model fitted using either `alldist` or `allvc` under null hypothesis.

`subset.formula` this is the subset part of the linear predictor under alternative hypothesis which does not include the null hypothesis linear predictor. This should be informed as R `formula`.

`null.values` the values of $\boldsymbol{\beta}_1^0$ as a vector.

`sign.level` the significance level which by default is `0.05`.

`analytic.var` if `TRUE` (default) the function will estimate the Fisher information using the analytic formulae for variance or by the last EM approximation if `FALSE`.

More information is described embedded in the code as comments after `#`.

```
wald.test <- function(fit.null,
                      subset.formula,
                      null.values = 0,
                      sign.level = .05,
                      analytic.var = TRUE){

  #

  # function respvar: Estimates the response variance
  source("respvar.R")
  match.fun(respvar)
  k         <- length(fit.null$masses)
  dname     <- fit.null$call$data
  frml      <- update(fit.null$formula,
                      formula(paste("~.+",subset.formula[-1])))
  fit.alt   <- update(fit.null,frml,plot.opt=0,verbose=FALSE)
  X         <- model.matrix(fit.alt)
  cnmX      <- attr(X,"dimnames")[[2]]
  cnmXnull  <- attr(fit.null$coefficients,"names")
  vrtst     <- cnmX[!cnmX%in%cnmXnull]
  mssttst   <- grep("MASS",vrtst)

  if((class(fit.null)=="glmmNPML")&(length(mssttst)>0)){
    vrtst <- vrtst[-mssttst]
  }
  if(length(null.values)==1&length(vrtst)>1){
    null.values <- rep(null.values,length(vrtst))
  }
  attr(null.values, "names") <- vrtst

  # partitions of the model matrix
  X1 <- cbind(X[,vrtst])
  X2 <- cbind(X[,cnmX%in%cnmXnull])

  if(any(null.values!=0)){
    fit.null <- update(fit.null,
                       offset=X1[1:nrow(fit.null$data),]%*%null.
```

```
                          values,
38                      plot.opt=0,verbose=FALSE)
39  }
40
41  fit.alt.glm <- fit.alt$lastglm
42  class(fit.alt.glm) = "glm"
43
44  # estimated quantities under alternative model
45  mu.alt  <- fit.alt.glm$fitted.values
46  eta.alt <- fit.alt.glm$family$linkfun(mu.alt)
47  V.alt   <- fit.alt.glm$family$variance(mu.alt)
48  dmu.alt <- fit.alt.glm$family$mu.eta(eta.alt)
49  phi.alt <- switch(fit.alt$family$family,
50                    gaussian         = fit.alt$sdev$sdev^-2,
51                    Gamma            = fit.alt$shape$shape,
52                    inverse.gaussian = fit.alt$shape$shape,
53                    poisson          = 1,
54                    binomial         = 1)
55  omg.alt  <- as.vector(fit.alt$post.prob)
56  dik.alt  <- phi.alt*dmu.alt*omg.alt/V.alt
57  Dm.alt   <- Diagonal(length(dik.alt),dik.alt)
58  estimate <- fit.alt$coefficients[vrtst]
59  # response variance estimation
60  vy.alt   <- respvar(m=fit.alt, exact=analytic.var)
61  if(length(vy.alt)==1){
62    vy.alt <- rep(vy.alt, length(fit.alt$fitted.values))
63  }
64  # partitioned Fisher information
65  if(class(fit.alt) == "glmmGQ"){
66    Upsln.alt <- kronecker(Matrix(1, ncol=k, nrow=k),
67                           Diagonal(length(vy.alt),vy.alt))
68  }
69  else{
70    if(class(fit.alt) == "glmmNPML"){
71      Upsln.alt <- Diagonal(length(vy.alt)*k,rep(vy.alt,k))
72    }else{
73      stop("the object 'fit.alt' should be either 'glmmGQ' or '
```

```
          glmmNPML'.")
74      }
75    }
76    Psi.alt <- Dm.alt%*%Upsln.alt%*%Dm.alt
77
78    # inverse of the partitioned Fisher information
79    invtX2PsiX2.alt <- solve(t(X2)%*%Psi.alt%*%X2)
80    C.alt <- invtX2PsiX2.alt%*%t(X2)%*%Psi.alt%*%X1
81    R.alt = X1 - X2%*%C.alt
82    tRPsiR.alt = t(R.alt)%*%Psi.alt%*%R.alt
83
84    # test statistic, df and p-value
85    sw         <- drop(t(estimate-null.values)%*%tRPsiR.alt%*%(
          estimate-null.values))
86    names(sw) <- "statistic"
87    df         <- length(vrtst)
88    names(df) <- "parameter"
89    pval       <- pchisq(sw, df=df, lower=FALSE)
90
91    mthd       <- list(paste("Wald test for GLMwRE",
92                             ifelse(analytic.var,
93                                    "(by analytic variance)",
94                                    "(by EM variance estimate)")),
95                       "\n",
96                       paste("null model:",
97                             deparse(fit.null$formula)),
98                       paste("alt. model:", deparse(frml)))
99
100   rval <- list("statistic" = sw, "parameter" = df,
101                "p.value" = pval, null.values = null.values,
102                method = mthd, data.name = dname)
103
104   class(rval) <- "GLMwRE.test"
105   return(rval)
106
107 }
```

## A.4 Rao test

The `rao.test` function computes the Rao test for GLMwRE and takes the arguments listed below.

`fit.null` the model fitted using either `alldist` or `allvc` under null hypothesis.

`subset.formula` this is the subset part of the linear predictor under alternative hypothesis which does not include the null hypothesis linear predictor. This should be informed as R `formula`.

`null.values` the values of $\beta_1^0$ as a vector.

`sign.level` the significance level which by default is `0.05`.

`analytic.var` if `TRUE` (default) the function will estimate the Fisher information using the analytic formulae for variance or by the last EM approximation if `FALSE`.

More information is described embedded in the code as comments after `#`.

```
1  rao.test <- function(fit.null,
2                       subset.formula,
3                       null.values = 0,
4                       sign.level = .05,
5                       analytic.var = TRUE){
6
7    #
8
9    # function respvar: Estimates the response variance
10   source("respvar.R")
11   match.fun(respvar)
12   k     <- length(fit.null$masses)
13   dname <- fit.null$call$data
14   frml  <- update(fit.null$formula,
15                   formula(paste("~.+",subset.formula[-1])))
16   X <- model.matrix(frml,data=npmlreg:::expand(fit.null$data,k))
17   if(class(fit.null)=="glmmNPML"){
18     X <- X[,-1]
```

```
19  }
20  cnmX     <- attr(X,"dimnames")[[2]]
21  cnmXnull <- attr(fit.null$coefficients,"names")
22  X <- cbind(X,model.matrix(fit.null)[,!cnmXnull%in%cnmX])
23  vrtst <- cnmX[!cnmX%in%cnmXnull]
24  if(!is.integer(grep("MASS",vrtst))){
25    vrtst <- vrtst[-grep("MASS",vrtst)]
26  }
27  if(length(null.values)==1&length(vrtst)>1){
28    null.values <- rep(null.values,length(vrtst))
29  }
30  attr(null.values, "names") <- vrtst
31
32  # partitions of the model matrix
33  X1 <- cbind(X[,vrtst])
34  X2 <- cbind(X[,cnmX%in%cnmXnull])
35
36  if(any(null.values!=0)){
37    fit.null <- update(fit.null,
38                       offset=X1[1:nrow(fit.null$data),]%*%null.
                              values,
39                       plot.opt=0,verbose=FALSE)
40  }
41  fit.null.glm <- fit.null$lastglm
42  class(fit.null.glm) = "glm"
43
44  # estimated model quantities under null hypothesis
45  mu.null  <- fit.null.glm$fitted.values
46  eta.null <- fit.null.glm$family$linkfun(mu.null)
47  V.null   <- fit.null.glm$family$variance(mu.null)
48  dmu.null <- fit.null.glm$family$mu.eta(eta.null)
49  phi.null <- switch(fit.null$family$family,
50                     gaussian         = fit.null$sdev$sdev^-2,
51                     Gamma            = fit.null$shape$shape,
52                     inverse.gaussian = fit.null$shape$shape,
53                     poisson          = 1,
54                     binomial         = 1)
```

```
55  omg.null <- as.vector(fit.null$post.prob)
56  dik.null <- phi.null*dmu.null*omg.null/V.null
57  Dm.null  <- Diagonal(length(dik.null),dik.null)
58  # score vector under null hypothesis
59  scr.null <- t(X1)%*%Dm.null%*%residuals(fit.null.glm,"response")
60  # response variance under null hypothesis
61  vy.null <- respvar(m=fit.null, exact=analytic.var)
62  if(length(vy.null)==1){
63    vy.null <- rep(vy.null, length(fit.null$fitted.values))
64  }
65  # partitioned Fisher information
66  if(class(fit.null) == "glmmGQ"){
67    Upsln.null <- kronecker(Matrix(1, ncol=k, nrow=k),
68                            Diagonal(length(vy.null),vy.null))
69  }
70  else{
71    if(class(fit.null) == "glmmNPML"){
72      Upsln.null <- Diagonal(length(vy.null)*k,rep(vy.null,k))
73    }else{
74      stop("the object 'fit.null' should be either 'glmmGQ' or '
          glmmNPML'.")
75    }
76  }
77  Psi.null         <- Dm.null%*%Upsln.null%*%Dm.null
78  invtX2PsiX2.null <- solve(t(X2)%*%Psi.null%*%X2)
79  C.null           <- invtX2PsiX2.null%*%t(X2)%*%Psi.null%*%X1
80  R.null           <- X1 - X2%*%C.null
81  invtRPsiR.null   <- solve(t(R.null)%*%Psi.null%*%R.null)
82
83  # test statistic, df and p-value
84  sr        <- drop(t(scr.null)%*%invtRPsiR.null%*%scr.null)
85  names(sr) <- "statistic"
86  df        <- length(vrtst)
87  names(df) <- "parameter"
88  pval      <- pchisq(sr, df=df, lower=FALSE)
89  mthd      <- list(paste("Rao test for GLMwRE",
90                          ifelse(analytic.var,
```

```
91                                      "(by analytic variance)",
92                                      "(by EM variance estimate)")),
93                      "\n",
94                      paste("null model:",
95                            deparse(fit.null$formula)),
96                      paste("alt. model:", deparse(frml)))
97
98    rval <- list("statistic" = sr, "parameter" = df,
99                 "p.value" = pval, null.values = null.values,
100                method = mthd, data.name = dname)
101   class(rval) <- "GLMwRE.test"
102   return(rval)
103
104   }
```

## A.5 Gradient test

The `gradient.test` function computes the gradient test for GLMwRE and takes the arguments listed below.

`fit.null` the model fitted using either `alldist` or `allvc` under null hypothesis.

`subset.formula` this is the subset part of the linear predictor under alternative hypothesis which does not include the null hypothesis linear predictor. This should be informed as R `formula`.

`null.values` the values of $\boldsymbol{\beta}_1^0$ as a vector.

`sign.level` the significance level which by default is `0.05`.

More information is described embedded in the code as comments after `#`.

```
1  gradient.test <- function(fit.null,
2                            subset.formula,
3                            null.values = 0,
4                            sign.level = .05){
5    #
6    dname     <- fit.null$call$data
```

```
7   frml      <- update(fit.null$formula,
8                       formula(paste("~.+",subset.formula[-1])))
9   fit.alt   <- update(fit.null,frml,plot.opt=0,verbose=FALSE)
10  X         <- model.matrix(fit.alt)
11  cnmX      <- attr(X,"dimnames")[[2]]
12  cnmXnull  <- attr(fit.null$coefficients,"names")
13  vrtst     <- cnmX[!cnmX%in%cnmXnull]
14  mssttst   <- grep("MASS",vrtst)
15  if((class(fit.null)=="glmmNPML")&(length(mssttst)>0)){
16    vrtst <- vrtst[-mssttst]
17  }
18  if(length(null.values)==1&length(vrtst)>1){
19    null.values <- rep(null.values,length(vrtst))
20  }
21  attr(null.values, "names") <- vrtst
22  X1        <- cbind(X[,vrtst])
23
24  if(any(null.values!=0)){
25    fit.null <- update(fit.null,
26                       offset=X1[1:nrow(fit.null$data),]%*%null.
                              values,
27                       plot.opt=0,verbose=FALSE)
28  }
29
30  fit.null.glm <- fit.null$lastglm
31  class(fit.null.glm) = "glm"
32
33  # model quantities under null hypothesis
34  mu.null  <- fit.null.glm$fitted.values
35  eta.null <- fit.null.glm$family$linkfun(mu.null)
36  V.null   <- fit.null.glm$family$variance(mu.null)
37  dmu.null <- fit.null.glm$family$mu.eta(eta.null)
38  phi.null <- switch(fit.null$family$family,
39                     gaussian         = fit.null$sdev$sdev^-2,
40                     Gamma            = fit.null$shape$shape,
41                     inverse.gaussian = fit.null$shape$shape,
42                     poisson          = 1,
```

```
43                      binomial         = 1)
44   omg.null  <- as.vector(fit.null$post.prob)
45   dik.null  <- phi.null*dmu.null*omg.null/V.null
46   Dm.null   <- Diagonal(length(dik.null),dik.null)
47   # score vector under null hypothesis
48   scr.null  <- t(X1)%*%Dm.null%*%residuals(fit.null.glm,"response")
49   estimate  <- fit.alt$coefficients[vrtst]
50   # test statistic, df and p-value
51   st        <- drop(t(scr.null)%*%(estimate - null.values))
52   names(st) <- "statistic"
53   df        <- length(vrtst)
54   names(df) <- "parameter"
55   pval      <- pchisq(st, df=df, lower=FALSE)
56   mthd      <- list("Gradient test for GLMwRE",
57                     "\n",
58                      paste("null model:",
59                            deparse(fit.null$formula)),
60                      paste("alt. model:", deparse(frml)))
61
62   rval <- list("statistic" = st, "parameter" = df,
63                "p.value" = pval, null.values = null.values,
64                estimate = estimate, method = mthd,
65                data.name = dname)
66   class(rval) <- "GLMwRE.test"
67   return(rval)
68
69 }
```

## A.6 Tests output

This function works "behind the curtain" to ensure a standard and user friendly output for the four test functions. It is not necessary to run it as it does automatically together with any of `lr.test`, `wald.test`, `rao.test` and `gradient.test`.

```
1 print.GLMwRE.test <- function(x, digits = getOption("digits"),
     prefix = "\t", ...)
2 {
```

```
3    cat("\n")
4    cat(strwrap(x$method, prefix = prefix), sep = "\n")
5    cat("\n")
6    out <- character()
7    if(!is.null(x$statistic))
8      out <- c(out, paste(names(x$statistic), "=",
9                          format(signif(x$statistic, max(1L, digits -
                               2L)))))
10   if(!is.null(x$parameter))
11     out <- c(out, paste(names(x$parameter), "=",
12                         format(signif(x$parameter, max(1L, digits -
                               2L)))))
13   if(!is.null(x$p.value)) {
14     fp1 <- format.pval(x$p.value, digits = max(1L, digits - 3L))
15     out <- c(out, paste("p-value",
16                         if(substr(fp1, 1L, 1L) == "<") fp1 else
                               paste("=",fp1)))
17   }
18   cat(strwrap(paste(out, collapse = ", ")), sep = "\n")
19   cat("\n")
20   invisible(x)
21 }
```

# Appendix B

# Variance estimation

In this Appendix we show the variance estimation formulae obtained for GQ models in Section 4.2 of Chapter 4.

## B.1 Gaussian quadrature case

Supose a generalised linear model with linear predictor

$$\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i \quad \text{for } i \text{ in } 1, \ldots, n,$$

where $z_i$ has standard normal distribution. As a result, for $i$ in $1, \ldots, n$, we have

$$
\begin{aligned}
\mathrm{E}[z_i] &= 0 \\
\mathrm{E}[z_i^2] &= 1 \\
\mathrm{E}[z_i^3] &= 0 \\
\mathrm{E}[z_i^4] &= 3 \\
\mathrm{E}[z_i^5] &= 0 \\
\mathrm{E}[z_i^6] &= 15 \\
\mathrm{E}[z_i^7] &= 0 \\
\mathrm{E}[z_i^8] &= 105 \\
\mathrm{E}[z_i^9] &= 0 \\
\mathrm{E}[z_i^{10}] &= 945 \\
\mathrm{Var}[z_i] &= 1 \\
\mathrm{Var}[z_i^2] &= \mathrm{E}[z_i^4] - \mathrm{E}^2[z_i^2] = 3 - 1 = 2 \\
\mathrm{Var}[z_i^3] &= \mathrm{E}[z_i^6] - \cancelto{0}{\mathrm{E}^2[z_i^3]} = 15 \\
\mathrm{Var}[z_i^4] &= \mathrm{E}[z_i^8] - \mathrm{E}^2[z_i^4] = 105 - 9^2 = 96 \\
\mathrm{Var}[z_i^5] &= \mathrm{E}[z_i^{10}] - \cancelto{0}{\mathrm{E}^2[z_i^5]} = 945
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}[z_i, z_i^2] &= \cancelto{0}{\mathrm{E}[z_i^3]} - \cancelto{0}{\mathrm{E}[z_i]}\cancelto{1}{\mathrm{E}[z_i^2]} = 0 \\
\mathrm{Cov}[z_i, z_i^3] &= \mathrm{E}[z_i^4] - \cancelto{0}{\mathrm{E}[z_i]}\cancelto{0}{\mathrm{E}[z_i^3]} = 3 \\
\mathrm{Cov}[z_i, z_i^4] &= \cancelto{0}{\mathrm{E}[z_i^5]} - \cancelto{0}{\mathrm{E}[z_i]}\cancelto{3}{\mathrm{E}[z_i^4]} = 0 \\
\mathrm{Cov}[z_i, z_i^5] &= \mathrm{E}[z_i^6] - \cancelto{0}{\mathrm{E}[z_i]}\cancelto{0}{\mathrm{E}[z_i^5]} = 15 \\
\mathrm{Cov}[z_i^2, z_i^3] &= \cancelto{0}{\mathrm{E}[z_i^5]} - \cancelto{1}{\mathrm{E}[z_i^2]}\cancelto{0}{\mathrm{E}[z_i^3]} = 0 \\
\mathrm{Cov}[z_i^2, z_i^4] &= \mathrm{E}[z_i^6] - \mathrm{E}[z_i^2]\mathrm{E}[z_i^4] = 15 - 3 = 12 \\
\mathrm{Cov}[z_i^2, z_i^5] &= \cancelto{0}{\mathrm{E}[z_i^7]} - \cancelto{1}{\mathrm{E}[z_i^2]}\cancelto{0}{\mathrm{E}[z_i^5]} = 0 \\
\mathrm{Cov}[z_i^3, z_i^4] &= \cancelto{0}{\mathrm{E}[z_i^7]} - \cancelto{0}{\mathrm{E}[z_i^3]}\cancelto{3}{\mathrm{E}[z_i^4]} = 0 \\
\mathrm{Cov}[z_i^3, z_i^5] &= \mathrm{E}[z_i^8] - \cancelto{0}{\mathrm{E}[z_i^3]}\cancelto{0}{\mathrm{E}[z_i^5]} = 0 \\
\mathrm{Cov}[z_i^4, z_i^5] &= \cancelto{0}{\mathrm{E}[z_i^9]} - \cancelto{3}{\mathrm{E}[z_i^4]}\cancelto{0}{\mathrm{E}[z_i^5]} = 0
\end{aligned}
$$

The response $y_i$ has mean $\mathrm{E}[y_i] = \mathrm{E}[\mathrm{E}[y_i|z_i]] = \mathrm{E}[\mu(z_i)] = \mu_i$ and variance

$$
\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\mathrm{Var}[y_i|z_i]] + \mathrm{Var}[\mathrm{E}[y_i|z_i]] \\
&= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \quad \text{for } i \text{ in } 1, \ldots, n,
\end{aligned}
$$

where $V(\mu)$ is the variance function and $\phi^{-1}$ the dispersion parameter.

### B.1.1 Gaussian response

Variance function: $V(\mu) = 1$

#### Identity link

- Link function: $g(\mu) = \mu = \eta$
- Inverse of the link function: $g^{-1}(\eta) = \eta = \mu$

$$\begin{aligned}\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\ &= \mathrm{E}[\phi^{-1}] + \mathrm{Var}[\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i] \\ &= \phi^{-1} + \sigma^2 \cancel{\mathrm{Var}[z_i]}^{1} \\ &= \phi^{-1} + \sigma^2, \text{ for } i \text{ in } 1, \ldots, n.\end{aligned}$$

**Log link**

- Link function: $g(\mu) = \log(\mu) = \eta$
- Inverse of the link function: $g^{-1}(\eta) = \exp(\eta) = \mu$

$$\begin{aligned}\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\ &= \mathrm{E}[\phi^{-1}] + \mathrm{Var}[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}] \\ &= \phi^{-1} + \mathrm{E}[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\})^2] - \mathrm{E}^2[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}] \\ &= \phi^{-1} + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{E}[\exp\{2\sigma z_i\}] - (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\mathrm{E}[\exp\{\sigma z_i\}])^2 \\ &= \phi^{-1} + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{M}_Z(2\sigma) - (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\mathrm{M}_Z(\sigma))^2 \\ &= \phi^{-1} + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\exp\{\cancel{4}^{2}\sigma^2/\cancel{2}\} - (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\exp\{\sigma^2/2\})^2 \\ &= \phi^{-1} + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma^2)\} - \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \sigma^2\} \\ &= \phi^{-1} + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \sigma^2\}(\exp\{\sigma^2\} - 1), \text{ for } i \text{ in } 1, \ldots, n.\end{aligned}$$

**Inverse link**

- Link function: $g(\mu) = 1/\mu = \eta$
- Inverse of the link function: $g^{-1}(\eta) = 1/\eta = \mu$

$$\begin{aligned}\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\ &= \mathrm{E}[\phi^{-1}] + \mathrm{Var}[(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)^{-1}],\end{aligned}$$

By Taylor expansion around 0, we have

$$\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i} \approx \frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}} - \frac{\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2} + \frac{\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} - \frac{\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}$$

Thus,

$$\begin{aligned}
\mathrm{Var}\left[\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i}\right] &\approx \mathrm{Var}\left[\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}}-\frac{\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2}+\frac{\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3}-\frac{\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}\right] \\
&= \frac{\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}\mathrm{Var}[z_i]+\frac{\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}\mathrm{Var}[z_i^2]+\frac{\sigma^6}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^8}\mathrm{Var}[z_i^3]- \\
&\quad -\frac{2\sigma^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5}\cancelto{0}{\mathrm{Cov}[z_i,z_i^2]}+\frac{2\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}\mathrm{Cov}[z_i,z_i^3]- \\
&\quad -\frac{2\sigma^5}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^7}\cancelto{0}{\mathrm{Cov}[z_i^2,z_i^3]} \\
&= \frac{\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{2\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}+\frac{15\sigma^6}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^8}+\frac{6\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6} \\
&= \frac{\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{8\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}+\frac{15\sigma^6}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^8}.
\end{aligned}$$

Therefore

$$\mathrm{Var}(y_i) \approx \phi^{-1}+\frac{\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{8\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}+\frac{15\sigma^6}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^8} \text{ for } i \text{ in } 1,\ldots,n.$$

### B.1.2 Gamma response

Variance function: $V(\mu)=\mu^2$

#### Identity link

- Link function: $g(\mu)=\mu=\eta$
- Inverse of the link function: $g^{-1}(\eta)=\eta=\mu$

$$\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))]+\mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\phi^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i)^2]+\mathrm{Var}[\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i] \\
&= \phi^{-1}[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2+2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma\cancelto{0}{\mathrm{E}[z_i]}+\sigma^2\cancelto{1}{\mathrm{E}[z_i^2]}]+\sigma^2\cancelto{1}{\mathrm{Var}[z_i]} \\
&= (\phi^{-1}+1)\sigma^2+\phi^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2, \text{ for } i \text{ in } 1,\ldots,n.
\end{aligned}$$

#### Log link

- Link function: $g(\mu)=\log(\mu)=\eta$
- Inverse of the link function: $g^{-1}(\eta)=\exp(\eta)=\mu$

$$
\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\phi^{-1}(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\})^2] + \mathrm{Var}[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}] \\
&= (\phi^{-1}+1)\mathrm{E}[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\})^2] - \mathrm{E}^2[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}] \\
&= (\phi^{-1}+1)\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{E}[\exp\{2\sigma z_i\}] - (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\mathrm{E}[\exp\{\sigma z_i\}])^2 \\
&= (\phi^{-1}+1)\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{M}_Z(2\sigma) - (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\mathrm{M}_Z(\sigma))^2 \\
&= (\phi^{-1}+1)\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\exp\{\cancel{4}^{2}\sigma^2/\cancel{2}\} - (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\exp\{\sigma^2/2\})^2 \\
&= (\phi^{-1}+1)\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma^2)\} - \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \sigma^2\} \\
&= \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \sigma^2\}[(\phi^{-1}+1)\exp\{\sigma^2\} - 1], \text{ for } i \text{ in } 1,\ldots,n.
\end{aligned}
$$

**Inverse link**

- Link function: $g(\mu) = 1/\mu = \eta$
- Inverse of the link function: $g^{-1}(\eta) = 1/\eta = \mu$

$$
\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\phi^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)^{-2}] + \mathrm{Var}[(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)^{-1}],
\end{aligned}
$$

By Taylor expansion around 0, we have

$$
\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i} \approx \frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}} - \frac{\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2} + \frac{\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} - \frac{\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4},
$$

and

$$
\left(\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i}\right)^2 \approx \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2} - \frac{2\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} + \frac{3\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4} - \frac{4\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5}.
$$

Thus

$$
\begin{aligned}
\mathrm{E}\left[\left(\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i}\right)^2\right] &\approx \mathrm{E}\left[\frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2} - \frac{2\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} + \frac{3\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4} - \frac{4\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5}\right] \\
&= \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2} - \frac{2\sigma\cancel{\mathrm{E}[z_i]}^{0}}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} + \frac{3\sigma^2\mathrm{E}[z_i^2]}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4} - \frac{4\sigma^3\cancel{\mathrm{E}[z_i^3]}^{0}}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5} \\
&= \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2} + \frac{3\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}.
\end{aligned}
$$

and, similarly to Gaussian response with inverse link

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i}\right] &\approx \mathrm{Var}\left[\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}}-\frac{\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2}+\frac{\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3}-\frac{\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}\right]\\ &= \frac{\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{8\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}+\frac{15\sigma^6}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^8}.\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{Var}(y_i) \approx \phi^{-1}&\left[\frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2}+\frac{3\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}\right]+\\ &+\frac{\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{8\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}+\frac{15\sigma^6}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^8},\ \text{for } i \text{ in } 1,\ldots,n.\end{aligned}$$

### B.1.3 Poisson response

Variance function: $V(\mu)=\mu$

Dispersion parameter: $\phi^{-1}=1$

#### Log link

- Link function: $g(\mu)=\log(\mu)=\eta$
- Inverse of the link function: $g^{-1}(\eta)=\exp(\eta)=\mu$

$$\begin{aligned}\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))]+\mathrm{Var}[\mu(z_i)]\\ &= \mathrm{E}[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i\})]+\mathrm{Var}[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i\}]\\ &= (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\mathrm{E}[(\exp\{\sigma z_i\})]+(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2\mathrm{Var}[\exp\{\sigma z_i\}]\\ &= (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\mathrm{M}_Z(\sigma)+(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2\{\mathrm{E}[(\exp\{\sigma z_i\})^2]+\mathrm{E}^2[\exp\{\sigma z_i\}]\}\\ &= (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\exp\{\sigma^2/2\}+(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2[\mathrm{M}_Z(2\sigma)+\mathrm{M}_Z^2(\sigma)]\\ &= (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\exp\{\sigma^2/2\}+(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2(\exp\{2\sigma^2\}+\exp\{\sigma^2\})\\ &= (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\exp\{\sigma^2/2\}+(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2\exp\{\sigma^2\}(\exp\{\sigma^2\}+1)\\ &= (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\exp\{\sigma^2/2\}[1+(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})\exp\{\sigma^2/2\}(\exp\{\sigma^2\}+1)]\\ &\quad \text{for } i \text{ in } 1,\ldots,n.\end{aligned}$$

#### Identity link

- Link function: $g(\mu)=\mu=\eta$

- Inverse of the link function: $g^{-1}(\eta) = \eta = \mu$

$$\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i] + \mathrm{Var}[\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i] \\
&= \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma\cancelto{0}{\mathrm{E}[z_i]} + \sigma^2\cancelto{1}{\mathrm{Var}[z_i]} \\
&= \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma^2 \quad \text{for } i \text{ in } 1, \ldots, n.
\end{aligned}$$

**Square root link**

- Link function: $g(\mu) = \sqrt{\mu} = \eta$
- Inverse of the link function: $g^{-1}(\eta) = \eta^2 = \mu$

$$\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)^2] + \mathrm{Var}[(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)^2] \\
&= \mathrm{E}[((\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma z_i + \sigma^2 z_i^2)] + \\
&\quad + \mathrm{Var}[((\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma z_i + \sigma^2 z_i^2)] \\
&= (\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma\cancelto{0}{\mathrm{E}[z_i]} + \sigma^2\cancelto{1}{\mathrm{E}[z_i^2]} + \\
&\quad + 4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^2\cancelto{1}{\mathrm{Var}[z_i]} + \sigma^4\mathrm{Var}[z_i^2] + 4(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^3\mathrm{Cov}[z, z_i^2] \\
&= (\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + \sigma^2 + 4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^2 + \sigma^4(\mathrm{E}[z_i^4] - \mathrm{E}^2[z_i^2]) + \\
&\quad + 4(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^3(\cancelto{0}{\mathrm{E}[z_i^3]} - \cancelto{0}{\mathrm{E}[z_i]}\cancelto{1}{\mathrm{E}[z_i^2]}) \\
&= (\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^2 + \sigma^2 + \sigma^4(3 - 1) \\
&= (\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^2 + \sigma^2 + 2\sigma^4, \text{ for } i \text{ in } 1, \ldots, n.
\end{aligned}$$

### B.1.4 Binomial response

Variance function: $V(\mu) = \mu(1 - \mu)$

Dispersion parameter: $\phi^{-1} = 1$

**Logit link**

- Link function: $g(\mu) = \log\left(\dfrac{\mu}{1-\mu}\right) = \eta$
- Inverse of the link function: $g^{-1}(\eta) = \dfrac{\exp(\eta)}{\exp(\eta)+1} = \mu$

$$
\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i) - \mu(z_i)^2] + \mathrm{E}[\mu(z_i)^2] - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i)] - \cancel{\mathrm{E}[\mu(z_i)^2]} + \cancel{\mathrm{E}[\mu(z_i)^2]} - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}\left[\frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\} + 1}\right] - \mathrm{E}^2\left[\frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\} + 1}\right].
\end{aligned}
$$

By Taylor expansion around 0, we have

$$
\begin{aligned}
\frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\} + 1} \approx \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1} &+ \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\sigma z_i}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^2} - \\
&- \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2 z_i^2}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^3} + \\
&+ \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^3 - 4(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 + \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^3 z_i^3}{6(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^4}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathrm{E}\left[\frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\} + 1}\right] \approx \mathrm{E}\Bigg[\frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1} &+ \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\sigma z_i}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^2} - \\
&- \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2 z_i^2}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^3} + \\
&+ \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^3 - 4(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 + \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^3 z_i^3}{6(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^4}\Bigg] \\
= \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1} &+ \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\sigma\cancelto{0}{\mathrm{E}[z_i]}}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^2} - \\
&- \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2 \mathrm{E}[z_i^2]}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^3} + \\
&+ \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^3 - 4(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 + \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^3\cancelto{0}{\mathrm{E}[z_i^3]}}{6(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^4} \\
= \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1} &- \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\} + 1)^3}.
\end{aligned}
$$

Therefore

$$\begin{aligned}
\mathrm{Var}(y_i) \approx & \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1} - \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^3} - \\
& - \left[\frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1} - \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^3}\right]^2 \\
= & \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1} - \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^3} - \\
& - \left[\frac{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^2} - \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^3 - (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2]\sigma^2}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^4} + \right. \\
& \left. + \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]^2\sigma^4}{4(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^6}\right] \\
= & \frac{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1} - \frac{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^2} - \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^3} + \\
& + \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^3 - (\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2]\sigma^2}{(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^4} - \\
& - \frac{[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]^2\sigma^4}{4(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}+1)^6}, \text{ for } i \text{ in } 1,\ldots,n.
\end{aligned}$$

**Probit link**

- Link function: $g(\mu) = \Phi^{-1}(\mu) = \eta$
- Inverse of the link function: $g^{-1}(\eta) = \Phi(\eta) = \mu$

$$\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i) - \mu(z_i)^2] + \mathrm{E}[\mu(z_i)^2] - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i)] - \cancel{\mathrm{E}[\mu(z_i)^2]} + \cancel{\mathrm{E}[\mu(z_i)^2]} - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}[\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)] - \mathrm{E}^2[\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)].
\end{aligned}$$

By Taylor expansion around 0, we have

$$\begin{aligned}
\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i) \approx \Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \sigma\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})z_i - \frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})z_i^2}{2} + \\
+ \frac{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]\sigma^3\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})z_i^3}{6},
\end{aligned}$$

where $\phi(\,\cdot\,)$ is the standard normal density. Thus,

$$\begin{aligned}
\mathrm{E}[\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i)] &\approx \mathrm{E}\left[\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})+\sigma\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})z_i-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})z_i^2}{2}+\right.\\
&\qquad\left.+\frac{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2+1]\sigma^3\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})z_i^3}{6}\right]\\
&= \Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})+\sigma\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})\cancelto{0}{\mathrm{E}[z_i]}-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})\cancelto{1}{\mathrm{E}[z_i^2]}}{2}+\\
&\quad+\frac{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2+1]\sigma^3\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})\cancelto{0}{\mathrm{E}[z_i^3]}}{6}\\
&= \Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{2},
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}^2[\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i)] &\approx \left[\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{2}\right]^2\\
&= \Phi^2(\boldsymbol{x}_i^\top\boldsymbol{\beta})-\cancel{2}\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{\cancel{2}}+\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^4\phi^2(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{4}\\
&= \Phi^2(\boldsymbol{x}_i^\top\boldsymbol{\beta})-(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})+\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^4\phi^2(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{4}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(y_i) \approx \Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{2}-\Phi^2(\boldsymbol{x}_i^\top\boldsymbol{\beta})+(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})\Phi(\boldsymbol{x}_i^\top\boldsymbol{\beta})-\\
-\frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma^4\phi^2(\boldsymbol{x}_i^\top\boldsymbol{\beta})}{4},\ \text{for } i \text{ in } 1,\ldots,n.
\end{aligned}$$

**Cauchit link**

- Link function: $g(\mu)=\tan\left[\pi\left(\mu-\frac{1}{2}\right)\right]=\eta$
- Inverse of the link function: $g^{-1}(\eta)=\frac{1}{\pi}\arctan(\eta)+\frac{1}{2}=\mu$

$$
\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i) - \mu(z_i)^2] + \mathrm{E}[\mu(z_i)^2] - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i)] - \cancel{\mathrm{E}[\mu(z_i)^2]} + \cancel{\mathrm{E}[\mu(z_i)^2]} - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}\left[\frac{1}{\pi}\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i) + \frac{1}{2}\right] - \mathrm{E}^2\left[\frac{1}{\pi}\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i) + \frac{1}{2}\right] \\
&= \frac{1}{\pi}\mathrm{E}[\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)] + \frac{1}{2} - \left(\frac{1}{\pi}\mathrm{E}[\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)] + \frac{1}{2}\right)^2 \\
&= \cancel{\frac{1}{\pi}\mathrm{E}[\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)]} + \frac{1}{2} - \\
&\quad - \left(\frac{1}{\pi^2}\mathrm{E}^2[\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)] + \cancel{\frac{1}{\pi}\mathrm{E}[\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)]} + \frac{1}{4}\right) \\
&= \frac{1}{4} - \frac{1}{\pi^2}\mathrm{E}^2[\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)].
\end{aligned}
$$

By Taylor expansion around 0, we have

$$
\begin{aligned}
\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i) \approx \arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \frac{\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1} - \frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2 z_i^2}{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]^2} + \\
+ \frac{[3(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 - 1]\sigma^3 z_i^3}{3[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]^3}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathrm{E}[\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)] &\approx \mathrm{E}\left[\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \frac{\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1} - \frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2 z_i^2}{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]^2} + \right. \\
&\quad \left. + \frac{[3(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 - 1]\sigma^3 z_i^3}{3[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]^3}\right] \\
&= \arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \frac{\sigma\cancelto{0}{\mathrm{E}[z_i]}}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1} - \frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\cancelto{1}{\mathrm{E}[z_i^2]}}{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]^2} + \\
&\quad + \frac{[3(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 - 1]\sigma^3\cancelto{0}{\mathrm{E}[z_i^3]}}{3[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]^3} \\
&= \arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta}) - \frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2}{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]^2}.
\end{aligned}
$$

Therefore,

$$
\mathrm{Var}(y_i) \approx \frac{1}{4} - \frac{1}{\pi^2}\left\{\arctan(\boldsymbol{x}_i^\top\boldsymbol{\beta}) - \frac{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2}{[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 + 1]^2}\right\}^2, \text{ for } i \text{ in } 1, \ldots, n.
$$

**Log link**

- Link function: $g(\mu) = \log(\mu) = \eta$

- Inverse of the link function: $g^{-1}(\eta) = \exp\{\eta\} = \mu$

$$\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i) - \mu(z_i)^2] + \mathrm{E}[\mu(z_i)^2] - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i)] - \cancel{\mathrm{E}[\mu(z_i)^2]} + \cancel{\mathrm{E}[\mu(z_i)^2]} - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}] - \mathrm{E}^2[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}] \\
&= \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\mathrm{E}[\exp\{\sigma z_i\}] - \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{E}^2[\exp\{\sigma z_i\}] \\
&= \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\mathrm{M}_Z(\sigma) - \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{M}_Z^2(\sigma) \\
&= \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\exp\left\{\frac{\sigma^2}{2}\right\} - \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\exp\{\sigma^2\} \\
&= \exp\left\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \frac{\sigma^2}{2}\right\} - \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \sigma^2\}, \text{ for } i \text{ in } 1, \ldots, n.
\end{aligned}$$

**Complementary log-log link**

- Link function: $g(\mu) = \log(-\log(1-\mu)) = \eta$
- Inverse of the link function: $g^{-1}(\eta) = 1 - \exp\{-\exp\{\eta\}\} = \mu$

$$\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i) - \mu(z_i)^2] + \mathrm{E}[\mu(z_i)^2] - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}[\mu(z_i)] - \cancel{\mathrm{E}[\mu(z_i)^2]} + \cancel{\mathrm{E}[\mu(z_i)^2]} - \mathrm{E}^2[\mu(z_i)] \\
&= \mathrm{E}[1 - \exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}\}] - \mathrm{E}^2[1 - \exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}\}] \\
&= \cancel{1} - \mathrm{E}[\exp\{\cancel{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}\}}] - \\
&\quad \cancel{1} + 2\mathrm{E}[\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}\}] - \mathrm{E}^2[\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}\}] \\
&= \mathrm{E}[\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}\}] - \mathrm{E}^2[\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}\}]
\end{aligned}$$

By Taylor expansion around 0, we have

$$\begin{aligned}
\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}\} \approx \exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}\sigma z_i + \\
+ \frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2 z_i^2}{2\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}} - \\
- \frac{[\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - 3\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} + \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^3 z_i^3}{6\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}}.
\end{aligned}$$

Thus,

$$
\begin{aligned}
\mathrm{E}[\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i\}\}] \approx \mathrm{E}\Big[&\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}\sigma z_i + \\
&+ \frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2 z_i^2}{2\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}} - \\
&- \frac{[\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - 3\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} + \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^3 z_i^3}{6\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}}\Big] \\
&= \exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}\sigma\cancelto{0}{\mathrm{E}[z_i]} + \\
&+ \frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2\cancelto{1}{\mathrm{E}[z_i^2]}}{2\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}} - \\
&- \frac{[\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - 3\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} + \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^3\cancelto{0}{\mathrm{E}[z_i^3]}}{6\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}} \\
&= \exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\} + \frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{Var}(y_i) &\approx \exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\} + \frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}} - \\
&- \left[\exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\} + \frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}}\right]^2 \\
&= \exp\{-\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\} + \frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{2\exp\{\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}} - \\
&- \exp\{-2\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\} - \frac{[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}]\sigma^2}{\exp\{2\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}} - \\
&- \frac{[\exp\{4(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} - 2\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\} + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}]\sigma^4}{4\exp\{2\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}\}\}},
\end{aligned}
$$

for $i$ in $1, \ldots, n$.

### B.1.5 Inverse gaussian response

Variance function: $V(\mu) = \mu^3$

**$1/\mu^2$ link**

- Link function: $g(\mu) = \dfrac{1}{\mu^2} = \eta$
- Inverse of the link function: $g^{-1}(\eta) = \dfrac{1}{\eta^{1/2}} = \mu$

$$\begin{aligned}\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\ &= \mathrm{E}[\phi^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)^{-3/2}] + \mathrm{Var}[(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)^{-1/2}].\end{aligned}$$

By Taylor expansion around 0, we have

$$\left(\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i}\right)^{1/2} \approx \frac{\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}{\boldsymbol{x}_i^\top\boldsymbol{\beta}} - \frac{\sigma z_i}{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} + \frac{3\sigma^2 z_i^2}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} - \frac{5\sigma^3 z_i^3}{16(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}},$$

and

$$\left(\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i}\right)^{3/2} \approx \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} - \frac{3\sigma z_i}{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} + \frac{15\sigma^2 z_i^2}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} - \frac{35\sigma^3 z_i^3}{16(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}.$$

Thus

$$\begin{aligned}\mathrm{Var}\left[\left(\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i}\right)^{1/2}\right] &\approx \mathrm{Var}\left[\frac{\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}{\boldsymbol{x}_i^\top\boldsymbol{\beta}} - \frac{\sigma z_i}{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} + \frac{3\sigma^2 z_i^2}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} - \right. \\ &\qquad \left. - \frac{5\sigma^3 z_i^3}{16(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}\right] \\ &= \frac{\sigma^2\mathrm{Var}[z_i]}{4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} + \frac{9\sigma^4\mathrm{Var}[z_i^2]}{64(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5} + \frac{25\sigma^6\mathrm{Var}[z_i^3]}{256(\boldsymbol{x}_i^\top\boldsymbol{\beta})^7} - \\ &\quad - \frac{3\sigma^3\mathrm{Cov}[z_i, z_i^2]}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4} + \frac{5\sigma^4\mathrm{Cov}[z, z_i^3]}{16(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5} - \frac{15\sigma^5\mathrm{Cov}[z_i^2, z_i^3]}{64(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6} \\ &= \frac{\sigma^2}{4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} + \frac{\sigma^4}{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5} + \frac{375\sigma^6}{256(\boldsymbol{x}_i^\top\boldsymbol{\beta})^7}.\end{aligned}$$

Moreover

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i}\right)^{3/2}\right] &\approx \mathrm{E}\left[\frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} - \frac{3\sigma z_i}{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} + \right.\\
&\qquad \left. + \frac{15\sigma^2 z_i^2}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} - \frac{35\sigma^3 z_i^3}{16(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}\right]\\
&= \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} - \frac{3\sigma\cancelto{0}{\mathrm{E}[z_i]}}{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} + \frac{15\sigma^2\cancelto{1}{\mathrm{E}[z_i^2]}}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} -\\
&\quad - \frac{35\sigma^3\cancelto{0}{\mathrm{E}[z_i^3]}}{16(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}\\
&= \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} + \frac{15\sigma^2}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{Var}(y_i) \approx \phi^{-1}&\left[\frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}} + \frac{15\sigma^2}{8(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top\boldsymbol{\beta}}}\right] +\\
&+ \frac{\sigma^2}{4(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} + \frac{\sigma^4}{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5} + \frac{375\sigma^6}{256(\boldsymbol{x}_i^\top\boldsymbol{\beta})^7},\ \text{for } i \text{ in } 1,\ldots,n.
\end{aligned}$$

**Inverse link**

- Link function: $g(\mu) = 1/\mu = \eta$
- Inverse of the link function: $g^{-1}(\eta) = 1/\eta = \mu$

$$\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)]\\
&= \mathrm{E}[\phi^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i)^{-3}] + \mathrm{Var}[(\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i)^{-1}],
\end{aligned}$$

By Taylor expansion around 0, we have

$$\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i} \approx \frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}} - \frac{\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2} + \frac{\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} - \frac{\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4},$$

and

$$\left(\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i}\right)^3 \approx \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3} - \frac{3\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4} + \frac{6\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5} - \frac{10\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}.$$

Thus

$$\begin{aligned}\mathrm{E}\left[\left(\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i}\right)^3\right] &\approx \mathrm{E}\left[\frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3}-\frac{3\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{6\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5}-\frac{10\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}\right]\\ &= \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3}-\frac{3\sigma\cancelto{0}{\mathrm{E}[z_i]}}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{6\sigma^2\cancelto{1}{\mathrm{E}[z_i^2]}}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5}-\frac{10\sigma^3\cancelto{0}{\mathrm{E}[z_i^3]}}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}\\ &= \frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3}+\frac{6\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5}.\end{aligned}$$

and, similarly to Gaussian response with inverse link

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i}\right] &\approx \mathrm{Var}\left[\frac{1}{\boldsymbol{x}_i^\top\boldsymbol{\beta}}-\frac{\sigma z_i}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2}+\frac{\sigma^2 z_i^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3}-\frac{\sigma^3 z_i^3}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}\right]\\ &= \frac{\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{8\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}+\frac{15\sigma^6}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^8}.\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{Var}(y_i) \approx \phi^{-1}&\left[\frac{1}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3}+\frac{6\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^5}\right]+\\ &+\frac{\sigma^2}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^4}+\frac{8\sigma^4}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^6}+\frac{15\sigma^6}{(\boldsymbol{x}_i^\top\boldsymbol{\beta})^8},\ \text{for } i \text{ in } 1,\ldots,n.\end{aligned}$$

**Identity link**

- Link function: $g(\mu)=\mu=\eta$
- Inverse of the link function: $g^{-1}(\eta)=\eta=\mu$

$$\begin{aligned}\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))]+\mathrm{Var}[\mu(z_i)]\\ &= \phi^{-1}\mathrm{E}[(\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i)^3]+\mathrm{Var}[\boldsymbol{x}_i^\top\boldsymbol{\beta}+\sigma z_i]\\ &= \phi^{-1}[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3+3(\boldsymbol{x}_i^\top\boldsymbol{\beta})^2\sigma\cancelto{0}{\mathrm{E}[z_i]}+3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2\cancelto{1}{\mathrm{E}[z_i^2]}+\sigma^3\cancelto{0}{\mathrm{E}[z_i^3]}+\\ &\quad+\sigma^2\cancelto{1}{\mathrm{Var}[z_i]}\\ &= \phi^{-1}[(\boldsymbol{x}_i^\top\boldsymbol{\beta})^3+3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\sigma^2]+\sigma^2,\ \text{for } i \text{ in } 1,\ldots,n.\end{aligned}$$

**Log link**

- Link function: $g(\mu)=\log(\mu)=\eta$
- Inverse of the link function: $g^{-1}(\eta)=\exp\{\eta\}=\mu$

$$
\begin{aligned}
\mathrm{Var}(y_i) &= \mathrm{E}[\phi^{-1}V(\mu(z_i))] + \mathrm{Var}[\mu(z_i)] \\
&= \phi^{-1}\mathrm{E}[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\})^3] + \mathrm{Var}[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}] \\
&= \phi^{-1}\mathrm{E}[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\})^3] + \mathrm{E}[(\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\})^2] - \\
&\quad - \mathrm{E}^2[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}] \\
&= \phi^{-1}\mathrm{E}[\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)\}] + \mathrm{E}[\exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i)\}] - \\
&\quad - (\mathrm{E}[\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma z_i\}])^2 \\
&= \phi^{-1}\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{E}[\exp\{3\sigma z_i\}] + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{E}[\exp\{2\sigma z_i\}] - \\
&\quad - \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{E}^2[\exp\{\sigma z_i\}] \\
&= \phi^{-1}\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\mathrm{M}_Z(3\sigma) + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}[\mathrm{M}_Z(2\sigma) - \mathrm{M}_Z^2(\sigma)] \\
&= \phi^{-1}\exp\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\exp\left\{\frac{9\sigma^2}{2}\right\} + \\
&\quad + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta})\}\left[\exp\left\{\frac{4\sigma^2}{2}\right\} - \exp\left\{\frac{\sigma^2}{2}\right\}\right] \\
&= \phi^{-1}\exp\left\{3(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \frac{9\sigma^2}{2}\right\} + \exp\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma^2)\} - \\
&\quad - \exp\left\{2(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + \frac{\sigma^2}{2}\right\}, \text{ for } i \text{ in } 1, \ldots, n.
\end{aligned}
$$